



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Académico Profesional de Estadística

**Modelo triangular de respuesta no aleatorizada: una
aplicación para estudiar el comportamiento de los
estudiantes de la FCM de la UNMSM frente a
preguntas sensibles**

TESIS

Para optar el Título Profesional de Licenciada en Estadística

AUTOR

Rosa Cristina RAMIREZ CARRILLO

ASESOR

Olga Lidia SOLANO DÁVILA

Lima, Perú

2016



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Ramirez, R. (2016). *Modelo triangular de respuesta no aleatorizada: una aplicación para estudiar el comportamiento de los estudiantes de la FCM de la UNMSM frente a preguntas sensibles*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Académico Profesional de Estadística]. Repositorio institucional Cybertesis UNMSM.

ESCUELA ACADÉMICO PROFESIONAL DE ESTADÍSTICA

ACTA DE SUSTENTACIÓN DE TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE
LICENCIADO EN ESTADÍSTICA

En la Ciudad Universitaria, Facultad de Ciencias Matemáticas, siendo las 17:15 horas del día 05 de Julio del año 2016 se reunieron los docentes designados como miembros del Jurado:

Lic. Félix Manuel Bartolo Gotarate

(Presidente)

Lic. Geraldine Judith Vigo Chacón

(Miembro)

Mg. Olga Lidia Solano Dávila

(Miembro Asesor)

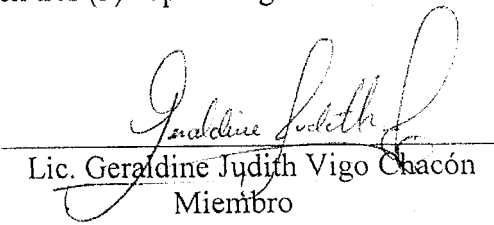
Para la sustentación de la Tesis intitulada "MODELO TRIANGULAR DE RESPUESTA NO ALEATORIZADA: UNA APLICACIÓN PARA ESTUDIAR EL COMPORTAMIENTO DE LOS ESTUDIANTES DE LA FCM DE LA UNMSM FRENTE A PREGUNTAS SENSIBLES", presentada por la Bachiller Rosa Cristina Ramirez Carrillo, para obtener el Título Profesional de Licenciada en Estadística.


Luego de la exposición de la Tesis, el Presidente invitó a la expositora a dar respuesta a las preguntas formuladas.

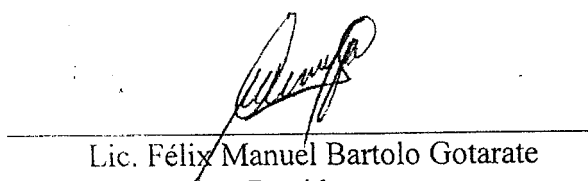
Realizada la evaluación correspondiente por los miembros del jurado, la expositora mereció la aprobación..... Sobresaliente con un calificativo promedio de Dieciocho / 18 (letras y números).

A continuación los miembros del jurado, dan manifiesto que la participante Bachiller Rosa Cristina Ramirez Carrillo, en virtud de haber aprobado la sustentación de su tesis, será propuesto para que se le otorgue el Título Profesional de Licenciada en Estadística.

Siendo las 18:00 horas, se levantó la Sesión, firmando para constancia la presente Acta en tres (3) copias originales.


Lic. Geraldine Judith Vigo Chacón
Miembro


Mg. Olga Lidia Solano Dávila
Miembro Asesor


Lic. Félix Manuel Bartolo Gotarate
Presidente

**MODELO TRIANGULAR DE RESPUESTA NO ALEATORIZADA:
UNA APLICACIÓN PARA ESTUDIAR EL COMPORTAMIENTO
DE LOS ESTUDIANTES DE LA FCM DE LA UNMSM FRENTE A
PREGUNTAS SENSIBLES**

BR. ROSA CRISTINA RAMIREZ CARRILLO

Tesis presentada a consideración del Cuerpo Docente de la Facultad de Ciencias Matemáticas, de la Universidad Nacional Mayor de San Marcos, como parte de los requisitos para obtener el Título Profesional de Licenciada en estadística.

Aprobada por:

Lic. Félix Manuel Bartolo Gotarate
(Presidente)

Lic. Geraldine Judith Vigo Chacón
(Miembro – Jurado)

Mg. Olga Lidia Solano Dávila
(Miembro – Asesor)

Lima – Perú

Julio 2016

FICHA CATALOGRÁFICA

RAMIREZ CARRILLO ROSA CRISTINA

Modelo triangular de respuesta no aleatorizada: una aplicación para estudiar el comportamiento de los estudiantes de la fcm de la unmsm frente a preguntas sensibles (Lima) 2016.

ix, 72 p, 29.7 cm (UNMSM, Licenciada, Estadística, 2016)

Tesis, Universidad Nacional Mayor de San Marcos.

Facultad de Ciencias Matemáticas 1. Estadística.

I. UNMSM/FACULTAD DE CIENCIAS MATEMÁTICAS

DEDICATORIA

A Dios por la bendición que permite hacer este
trabajo posible.

A mi madre por su amor y dedicación.

A mi padre por ser el hombre de mi vida.

A Hassan por la amistad y el amor que nos une.

A Stefany, Nathalie y Sabrina por enseñarme el
valor de la amistad a pesar de la distancia y el
tiempo.

AGRADECIMIENTOS

A mi tutora, la Magister Olga Lidia Solano Dávila, por su paciencia, por hacer que este trabajo sea hecho con esfuerzo y buenos momentos.

Gracias por su dedicación y el valioso aporte que hace hoy en mi vida profesional. De igual manera el agradecimiento a todos mis profesores que estuvieron presentes en mi formación profesional.

RESUMEN

Modelo triangular de respuesta no aleatorizada: Una aplicación para estudiar el comportamiento de los estudiantes de la FCM de la UNMSM frente a preguntas sensibles

BR. Rosa Cristina Ramirez Carrillo

Julio – 2016

ASESORA : Mg. Olga Lidia Solano Dávila

TÍTULO OBTENIDO : Licenciada en Estadística

El modelo triangular de respuesta no aleatorizada (MTRNA) propuesto por Yu, Tian y Tang (Yu, Tian, & Tang, 2008); fue aplicado en un estudio realizado en la Facultad de Ciencias Matemáticas, para investigar el comportamiento de los estudiantes con respecto a la copia en los exámenes, al uso de notas en teléfono móvil o calculadora durante un examen, uso de drogas para mejorar el rendimiento académico y la presentación de trabajos copiados; estos resultados se compararán con el método convencional de entrevista directa. La población objeto de estudio incluye a los estudiantes matriculados en el semestre 2015-I.

PALABRAS CLAVE:

**MODELO TRIANGULAR DE RESPUESTA NO ALEATORIZADA (MTRNA),
PREGUNTA SENSIBLE, ENTREVISTA DIRECTA, POTENCIA DE LA PRUEBA.**

ABSTRACT

Triangular Model of nonrandomized responds: An application to study the behavior of students of the Mathematics Department at National University of San Marcos facing sensitive questing.

BR. Rosa Cristina Ramirez Carrillo

July – 2016

ADVISOR : Mg. Olga Lidia Solano Dávila

DEGREE OBTAINED : Bachelor's degree in Statistics

Triangular Model of Nonrandomized Response for proposed Yu, Tian and Tang (Yu et al., 2008); was applied in a study in the Faculty of Mathematics Sciences, to investigate the behavior of students regarding copying in exams, the use of notes on mobile phone or calculator during an exam, use of drugs to improve academic performance and copy in the works; these results will be compared with the conventional method of direct interview. The study population included students enrolled in 2015-I .

KEYS WORDS:

RESPONSE TRIANGULAR MODEL NONRANDOMIZED, SENSIBLE QUESTION, INTERVIEW DIRECT, POWER TEST

ÍNDICE

CAPÍTULO I: MARCO REFERENCIAL	3
1.1. PLANTEAMIENTO DEL PROBLEMA.....	4
1.1.1. Objetivo general	4
1.1.2. Objetivos específicos.....	4
1.2. ANTECEDENTES DEL PROBLEMA.....	4
1.2.1. Nivel nacional	4
1.2.2. Nivel internacional	4
1.3. CONCEPTOS BÁSICOS	6
CAPÍTULO II: MODELO DE RESPUESTA NO ALEATORIZADA	7
2.1. LOS MODELOS DE RESPUESTA NO ALEATORIZADA	8
2.2. EL DISEÑO DE ENCUESTA PARA EL MODELO TRIANGULAR	8
2.2.1. Diseño del cuestionario.....	9
2.2.2. Análisis de precisión y poder.....	10
2.2.3. Condiciones	11
2.2.4. Varianza del estimador	11
2.2.5. Eficiencia relativa.....	13
2.2.6. Grado de protección de la privacidad del entrevistado.....	14
2.3. COMPARACIÓN DEL MODELO TRIANGULAR CON EL MODELO DE RESPUESTA ALEATORIZADA DE WARNER.....	15
2.3.1. La diferencia de dos varianzas.....	15
2.3.2. Eficiencia relativa del modelo de Warner y el modelo triangular	17
2.3.3. Grado de protección de la privacidad.....	18
2.4. PROPIEDADES ASINTÓTICAS DEL ESTIMADOR DE MÁXIMA VEROSIMILITUD.	19
2.4.1. Una derivación alternativa del EMV	19
2.4.2. Intervalo de confianza de Wald	20
2.5. EL MODELO TRIANGULAR PARA EL PROBLEMA DE UNA MUESTRA	20
2.5.1. Prueba de hipótesis unilateral.....	21
2.5.2. Prueba de hipótesis bilateral.....	22
2.5.3. Evaluación del desempeño comparando la potencia exacta y el poder asintótico.....	23
2.5.4. Evaluación del desempeño calculando nT con nT/nD	24
2.6. EL MODELO TRIANGULAR PARA EL PROBLEMA DE DOS MUESTRAS. .	26

CAPÍTULO III: APLICACIÓN DEL MTRNA	29
3.1. IMPLEMENTACIÓN DEL MODELO TRIANGULAR DE RESPUESTA NO ALEATORIZADA.....	30
3.1.1. Elección de las variables.....	30
3.1.2. Tamaño de muestra:	32
3.1.3. Plan de muestreo de la aplicación.	35
3.2. RESULTADOS DE LA APLICACIÓN DEL MODELO TRIANGULAR DE RESPUESTA NO ALEATORIZADA.....	37
3.2.1. Evaluación del rendimiento de la muestra.	37
3.2.2. Características de la población investigada	37
3.2.3. Resultados numéricos.....	37
3.3. IMPLEMENTACIÓN DEL MODELO DE ENTREVISTA DIRECTA.	40
3.3.1. Plan de muestreo de la aplicación.	41
3.4. RESULTADOS DE LA APLICACIÓN DEL MODELO DE ENTREVISTA DIRECTA.	43
3.4.1. Evaluación del rendimiento de la muestra.	43
3.4.2. Características de la población investigada	43
3.4.3. Resultados numéricos.....	43
3.5. CONCLUSIONES Y RECOMENDACIONES	48
ANEXOS	49
REFERENCIAS BIBLIOGRÁFICAS	71

INTRODUCCIÓN

Los modelos de respuesta no aleatorizada (MRNA) se han desarrollado recientemente para el análisis de cuestiones sensibles. A diferencia de los modelos de respuesta aleatoria tradicionales, los MRNA no requieren de un dispositivo de aleatorización.

Es conocido que la información sensible es buscada constantemente en estudios de comportamientos sociales con encuestas por muestreo que contienen preguntas sensibles como el comportamiento sexual y el aborto, el virus de la inmunodeficiencia humana (VIH), comportamientos ilegales como el consumo de drogas, el abuso, la violencia, y los ingresos personales.

Las desventajas de estos estudios se presentan cuando algunos encuestados pueden negarse a responder y aún peor dar respuestas falsas para mantener su privacidad cuando estas preguntas se hacen directamente. Además, la sensibilidad puede ser aguda en encuestas de diversas poblaciones. Por lo que los datos finales inevitablemente pueden incluir sesgo de respuesta. Como resultado, es difícil hacer inferencias válidas sobre la base de estos datos inexactos.

Por otro lado, los MRNA introducen una pregunta no sensible con los resultados binarios en el cuestionario con el fin de proteger la privacidad y fomentar la cooperación de los encuestados. Por desgracia, se asume la proporción de sujetos que poseen la característica no sensible debe ser conocido, y que la pregunta no sensible guarda una relación de independencia con la pregunta sensible.

Este trabajo tiene como objetivo mostrar las bondades del Modelo Triangular de Respuesta No aleatorizada (MTRNA) con respecto al modelo tradicional de entrevista directa; disminuyendo la no respuesta y el sesgo de respuesta en la encuesta con preguntas sensibles y/o delicadas. Proponemos el MTRNA que disminuyen los supuestos antes mencionados.

Este trabajo de investigación comprende 3 capítulos, donde en el Capítulo I se presenta el planteamiento del problema, los objetivos, antecedentes y conceptos básicos que serán útiles para una mejor comprensión del MTRNA.

El capítulo II contiene el marco teórico del MTRNA donde se desarrolla las estimaciones de los parámetros y de intervalos de confianza para la proporción sensible. Se investigarán las asignaciones del tamaño de muestra óptimo.

Por último para ilustrar la metodología propuesta el capítulo III presenta una aplicación del MTRNA realizada a los estudiantes de la FCM que están matriculados en el semestre 2015- I, así como los resultados y una comparación con el método tradicional (entrevista directa).

CAPÍTULO I: MARCO REFERENCIAL

1.1. PLANTEAMIENTO DEL PROBLEMA

1.1.1. Objetivo general

- Estudiar con detalle la teoría del Modelo Triangular de Respuesta No Aleatorizada y aplicarlo a nuestra realidad.

1.1.2. Objetivos específicos

- Conocer las ventajas y desventajas del Modelos Triangular de Respuesta No Aleatorizada.
- Aplicar el Modelo Triangular de Respuesta No Aleatorizada a los estudiantes de la Facultad de Ciencias Matemáticas matriculados en el semestre 2015-I, para estudiar el comportamiento de los alumnos frente a preguntas sensibles.
- Aplicar el método convencional a los estudiantes de la Facultad de Ciencias Matemáticas matriculados en el semestre 2015-I, para estudiar el comportamiento de los alumnos frente a preguntas sensibles.
- Comparar los resultados entre el MTRNA y la entrevista directa.

1.2. ANTECEDENTES DEL PROBLEMA

1.2.1. Nivel nacional

No se han encontrado investigaciones del modelo triangular de respuesta no aleatorizada en la UNMSM ni a nivel nacional.

1.2.2. Nivel internacional

A nivel internacional encontramos investigaciones sobre el MRNA sin que este sea necesariamente específico del modelo triangular, aquí mencionaremos algunas de estas:

Yu, Tian y Tang en el año 2008 propone el diseño triangular en uno de sus artículos para las encuestas que hacen preguntas sensibles con los resultados binarios. Sea $Y = 1$ que denota la clase de personas que poseen una característica sensible (por ejemplo, el consumo de drogas) y $Y = 0$ denota la clase complementaria. Sea W una variable dicotómica no sensible que es independiente de Y . Por ejemplo, $W = 1$ si el entrevistado

nació entre agosto y diciembre y $W = 0$ de otra forma. W debe elegirse convenientemente de modo que la proporción $p = P(W = 1)$ se conoce o se puede estimar fácilmente (Yu et al., 2008).

En el año 2009 los autores Ming T. Tan, Guo-Liang Tian y Man-Lai Tang en su artículo "Surveys with Sensitive Questions: A Nonrandomized Response Approach" (Tan, Tian, & Tang, 2009) presenta brevemente el enfoque de respuesta no aleatorizada (RNA), proponiendo varios modelos nuevos RNA, donde se compara la eficiencia de los modelos RNA y de los modelos de respuesta aleatorizada (RA) y estudia la viabilidad de los modelos RNA. En adición, se presenta el concepto del grado de protección de la privacidad entre el modelo RNA y el modelo de Warner para reflejar la medida en que la privacidad está protegida. Estos estudios muestran que no sólo el enfoque RNA está libre de las limitaciones del enfoque aleatorio, sino también que el modelo RNA en realidad aumenta la eficiencia relativa y el grado de protección de la privacidad. Por lo tanto, el enfoque de respuesta no aleatorizada ofrece una alternativa atractiva para el enfoque de respuesta aleatorizada.

En el año 2014, en el Libro "*Incomplete Categorical Data Design*", los autores Tian y Tang mencionan que las encuestas que buscan medir a partir de preguntas directas, características sensibles (aborto, fraude, consumo de drogas, relaciones sexuales, copia de exámenes etc.), suelen tropezar con dificultades como que los individuos no contesten o que intencionadamente falsifiquen sus respuestas. Existen múltiples estudios realizados para mejorar la calidad y veracidad de las respuestas obtenidas sobre temas sensibles. (Tian & Tang, 2014).

Frecuentemente los investigadores obtienen resistencias a la participación en estos estudios u obtienen respuestas falsas de parte de los entrevistados. Warner en 1965 realizó la primera propuesta para obtener respuestas válidas ante preguntas embarazosas basándose en la realización de dos preguntas mutuamente excluyentes (por ejemplo "A: Declaré mis ingresos extraordinarios el año pasado"; B: No declaré mis ingresos extraordinarios el año pasado"). (Basulto, J., 1982)

Posteriormente, se desarrollaron otros métodos basados en el modelo de Warner, como el denominado "método de alternativa forzada" de Tracy (Tracy PE & Fox JA, 1986). El Modelo de Respuesta no Aleatorizada es un método especialmente diseñado para

asegurar privacidad a los entrevistados en el estudio de temas sensibles, delicados o embarazosos. Se intenta con ello evitar el sesgo de los entrevistados en ciertas conductas hacia la respuesta socialmente más deseable. Se ha utilizado para analizar temas como copiar en los exámenes, fraudes, haber sido arrestado, conducir bajo los efectos del alcohol, tener un hijo fuera del matrimonio, aborto, etc.

1.3. CONCEPTOS BÁSICOS

RESPUESTA NO ALEATORIZADA.- Es la respuesta a la pregunta de interés, utilizando un mecanismo no aleatorio.

PREGUNTA DIRECTA.- Cuando el entrevistador hace la pregunta personalmente (de persona a persona).

PREGUNTA DELICADA O SENSIBLE.- Son aquellas preguntas comprometedoras, por ejemplo: ¿Usted es portador del VIH?, ¿Usas drogas?, etc.

PREGUNTA NO RELACIONADA.- Son aquellas preguntas que no tienen relación con la pregunta de interés, pero que sin embargo influyen para que la respuesta a la pregunta delicada sea verdadera. Por ejemplo, ¿Consumes drogas? (pregunta delicada), ¿Es Ud. mayor de 30 años? (pregunta no relacionada).

NO RESPUESTA.- Cuando la persona o entidad que se trata de entrevistar se rehúsa a colaborar, ya sea por razones personales o subjetivas. El problema es más frecuente cuando el cuestionario hace referencia a datos de conducta íntima o personal (gasto en drogas, abortos, gasto en bebidas alcohólicas, etc.) especialmente si el entrevistado tiene la creencia de que su respuesta puede perjudicarlo.

TASA DE RESPUESTA.- Porcentaje que resulta de dividir el total de respuestas obtenidas entre el total de encuestas realizadas en un determinado estudio, multiplicado por 100.

SESGO DE RESPUESTA.- Diferencia entre la respuesta esperada y la respuesta obtenida.

CAPÍTULO II: MODELO DE RESPUESTA NO ALEATORIZADA

2.1. LOS MODELOS DE RESPUESTA NO ALEATORIZADA

La adquisición de información sensible es a menudo necesaria en una amplia gama de aplicaciones estadísticas. Por ejemplo, es posible que algunos estudios conductuales y sociales para solicitar información sobre la historia reproductiva, el comportamiento sexual, uso de drogas ilegales, la violencia familiar y los ingresos. Cuando se pregunta directamente a las personas de la encuesta este tipo de preguntas sensibles, algunos suelen negarse a contestar o pueden dar respuestas falsas para proteger su privacidad. El problema es aún más complicado con las encuestas en diversas poblaciones, debido a la interacción de la sensibilidad y diversidad demandada. Por tanto, es difícil sacar conclusiones válidas de estos datos inexactos, ya que incluyen el sesgo de rechazo, el sesgo de respuesta, y tal vez ambas cosas. Ha sido durante mucho tiempo un desafío para obtener dicha información al tiempo que la privacidad del encuestado protegido y los datos resultantes se analizaron correctamente.

Mientras que el modelo de respuesta aleatorizada presenta limitaciones, por ejemplo falta de reproducibilidad y la confianza de los entrevistados así como un mayor costo debido a la utilización de dispositivos de aleatorización. La evolución reciente del enfoque de respuesta no aleatorizado tienen como objetivo aliviar o eliminar tales limitaciones. (Tan et al., 2009)

2.2. EL DISEÑO DE ENCUESTA PARA EL MODELO TRIANGULAR

Sea $\{Y = 1\}$ la clase de personas que poseen una característica sensible (por ejemplo, el consumo de drogas, hurto, evasión de impuestos, y así sucesivamente) y $\{Y = 0\}$ denota la clase complementaria. Sea W una variable no sensible dicotómica e independiente de Y . Es importante que los entrevistadores seleccionen una variable W adecuada de manera que la proporción $p = P(W = 1)$ se puede estimar fácilmente.

Entonces, se supone que p es conocido. Por ejemplo, podemos definir $W = 1$ si el encuestado nació entre agosto y diciembre y $W = 0$ en caso contrario. Por lo tanto, es razonable asumir que $p \approx 5/12 = 0.41667$. Nuestro objetivo es estimar la proporción:

$$\pi = P(Y = 1).$$

2.2.1. Diseño del cuestionario

Supongamos que a los investigadores les gustaría estimar la proporción de consumidores de drogas en una población específica. El objetivo podría lograrse a través de la recopilación de la información de mezcla de una pregunta sensible (por ejemplo, ¿es usted un usuario de drogas?) Y una pregunta no sensible no relacionada (por ejemplo, ¿su cumpleaños es entre agosto y diciembre?). Los entrevistadores pueden diseñar un cuestionario en el formato que se muestra en el lado izquierdo de la Tabla 2.1 y pedir a cada entrevistado poner una cruz en el círculo (es decir, $\{ Y = 0, W = 0 \}$) si él / ella pertenece a este círculo, caso contrario pone una cruz en el cuadrado superior (es decir , $\{ Y = 0, W = 1 \}$) si él / ella pertenece a uno de los tres cuadrados. Tenga en cuenta que tanto $\{ Y = 0, W = 0 \}$ y $\{ Y = 0, W = 1 \}$ son clases no sensibles, y la clase sensible $\{ Y = 1 \}$ se mezcla con otra subclase no sensible $\{ Y = 0, W = 1 \}$. Por lo tanto, un encuestado que es un consumidor de drogas puede proteger su verdadera identidad por aquellos usuarios no consumidores de drogas que nacen entre agosto y diciembre, y está, posiblemente, dispuesto a poner una cruz en el cuadrado superior. Tal diseño alienta a los encuestados de no sólo participar en la encuesta, sino también proporcionar respuestas verdaderas. Las probabilidades de las celdas correspondientes se enumeran en la parte derecha de la Tabla 2.1. Yu, Tian y Tang (2008) llamaron a este un modelo triangular, que pertenece a la clase de la admisibilidad de diseño definido por Nayak (Nayak, 1994).

Tabla 2.1: El modelo triangular y sus correspondientes celdas de probabilidades.

Categoría	W=0	W=1	Categoría	W=0	W=1	Marginal
Y=0	<input type="radio"/>	<input type="checkbox"/>	Y=0	$(1 - \pi)(1 - p)$	$(1 - \pi)p$	$1 - \pi$
Y=1	<input type="checkbox"/>	<input type="checkbox"/>	Y=1	$\pi(1 - p)$	πp	π
			Marginal	$(1 - p)$	p	1

Fuente: Adaptado de la Tabla 1 de Yu, Tian y Tang (2008).

Nota: Por favor, ponga una cruz en el círculo si usted pertenece a este círculo o poner una cruz en el cuadrado superior, si usted pertenece a uno de los tres cuadrados.

Una ventaja inmediata con el modelo triangular es su robustez a la falta de respuesta de la pregunta no sensible en el sentido de que permite tal falta de respuesta. Por ejemplo, para n encuestados, observamos n_1 cruces en el círculo, n_2 cruces en la cuadrado, y n_3 no

respuestas ($n = n_1 + n_2 + n_3$) Este resultado es equivalente a la observación $Y_{Obs} = \{n; n_1, n_2 + n_3\}$ (es decir, se observó n_1 cruces en el círculo y $n_2 + n_3$ cruces en el cuadrado superior) bajo el supuesto de que el encuestado está siempre dispuesto a responder a la pregunta si él / ella no es un consumidor de drogas.

2.2.2. Análisis de precisión y poder.

Uno de los pasos más importantes en las encuestas por muestreo es la determinación de la cantidad de participantes. En la práctica, los tamaños de muestra puede determinarse basándose en el método de análisis de la precisión o el método de análisis de potencia (Shein-Chung Chow, Hansheng Wang, & Jun Shao, 2003), que está estrechamente relacionada con la noción de error tipo I, tipo II y funciones de poder en una prueba de hipótesis.

Error tipo I, error tipo II y poder:

- (1) El rechazo de la hipótesis nula H_0 cuando esta es verdadera se llama error de tipo I. La probabilidad de cometer el error de tipo I:

$$\begin{aligned}\alpha &= P(\text{Error tipo I}) \\ &= P(\text{Rechaza } H_0 | H_0 \text{ es verdadera})\end{aligned}$$

- (2) La aceptación de la hipótesis nula H_0 cuando es falsa se llama error de tipo II. La probabilidad de cometer el error de tipo II:

$$\begin{aligned}\beta &= P(\text{Error tipo II}) \\ &= P(\text{Acepta } H_0 | H_0 \text{ es falsa})\end{aligned}$$

- (3) La potencia de la prueba es definida como la probabilidad de rechazar H_0 correctamente cuando la H_0 es falsa; esto es,

$$\begin{aligned}\text{Potencia} &= 1 - \beta \\ &= P(\text{Rechazar } H_0 | H_0 \text{ es falsa})\end{aligned}$$

2.2.3. Condiciones

- a) La tabla de contingencia debe presentar una pregunta sensible.
- b) La probabilidad de la pregunta no sensible (p) debe ser conocida.
- c) Para la evaluación del tamaño de muestra se debe considerar dos probabilidades hipotéticas de la pregunta sensible, esto debe ser elegido por el investigador.

2.2.4. Varianza del estimador

Para el modelo triangular en la tabla 2.1, se define una variable aleatoria de Bernoulli Y_T , como:

$$Y_T = \begin{cases} 1, & \text{si pertenece a cualquiera de los tres cuadrados.} \\ 0, & \text{si la cruz está en el círculo.} \end{cases} \quad (2.1)$$

Donde T representa el modelo triangular, las probabilidades $Y_T = 1$ y $Y_T = 0$ están dadas por:

$$P(Y_T = 1) = \pi + (1 - \pi)p$$

$$P(Y_T = 0) = (1 - \pi)(1 - p)$$

Donde p se asume como conocido. Sea $Y_{Obs} = \{y_{i,T} : i = 1, \dots, n\}$ los datos observados por los n encuestados con $y_{i,T} = 1$ si el i -ésimo encuestado pone una cruz en el cuadrado superior y $y_{i,T} = 0$ si el i -ésimo encuestado pone una cruz en el círculo. La función verosimilitud de π es:

$$L_T(\pi | Y_{Obs}) = \prod_{i=1}^n \{\pi + (1 - \pi)p\}^{y_{i,T}} \{(1 - \pi)(1 - p)\}^{1 - y_{i,T}} \quad (2.2)$$

El estimador de máxima verosimilitud de π y su varianza correspondiente están dadas por: (Ver Anexo 1)

$$\hat{\pi}_T = \frac{\bar{y}_T - p}{1 - p} \quad (2.3)$$

Donde $\bar{y}_T = \frac{1}{n} \sum_{i=1}^n y_{i,T}$, verificando que $\hat{\pi}_T$ es un estimador insesgado de π :

Prueba:

$$\begin{aligned}
E[\hat{\pi}_T] &= E\left[\frac{\bar{y}_t - p}{1 - p}\right] = E\left[\frac{\frac{1}{n}\sum_{i=1}^n y_{i,T} - p}{1 - p}\right] \\
&= E\left[\frac{\sum_{i=1}^n y_{i,T} - np}{n(1 - p)}\right] = \frac{E[\sum_{i=1}^n y_{i,T}] - np}{n(1 - p)} \\
&= \frac{n(\pi + (1 - \pi)p) - np}{n(1 - p)} = \frac{n\pi + np - n\pi p - np}{n(1 - p)} \\
&= \frac{n\pi + np - n\pi p - np}{n(1 - p)} = \frac{n\pi(1 - p)}{n(1 - p)} = \pi
\end{aligned}$$

Y la varianza de $\hat{\pi}_T$ está dada por:

$$Var(\hat{\pi}_T) = Var(\hat{\pi}_D) + \frac{p(1 - \pi)}{n(1 - p)}$$

Prueba:

$$Var(\hat{\pi}_T) = Var\left(\frac{\bar{y}_t - p}{1 - p}\right) = Var\left(\frac{\frac{1}{n}\sum_{i=1}^n y_{i,T} - p}{1 - p}\right) = \frac{Var(\sum_{i=1}^n y_{i,T})}{[n(1 - p)]^2}$$

Como $y_{i,T}$ es variable aleatoria con distribución Bernoulli, se tiene que

$$\begin{aligned}
Var(\hat{\pi}_T) &= \frac{n(\pi + (1 - \pi)p)(1 - \pi)(1 - p)}{[n^2(1 - p)^2]} = \frac{(\pi + (1 - \pi)p)(1 - \pi)}{[n(1 - p)]} \\
&= \frac{(\pi(1 - \pi) + (1 - \pi)p - \pi p(1 - \pi))}{[n(1 - p)]} \\
&= \frac{(\pi(1 - \pi)(1 - p) + (1 - \pi)p)}{[n(1 - p)]} = \frac{(\pi(1 - \pi)(1 - p))}{[n(1 - p)]} + \frac{p(1 - \pi)}{n(1 - p)} \\
Var(\hat{\pi}_T) &= \frac{\pi(1 - \pi)}{n} + \frac{p(1 - \pi)}{n(1 - p)}
\end{aligned}$$

La varianza con el modelo de entrevista directa $Var(\hat{\pi}_D)$ es definido como:

$$Var(\hat{\pi}_D) = \frac{\pi(1 - \pi)}{n} \tag{2.4}$$

$$Var(\hat{\pi}_T) = Var(\hat{\pi}_D) + \frac{p(1-\pi)}{n(1-p)} \quad (2.5)$$

Para cualquier π fijo, notamos que:

$$nVar(\hat{\pi}_T) = (1-\pi) \left(\pi + \frac{p}{1-p} \right) \quad (2.6)$$

Es una función creciente de p cuando $0 < p < 1$ y enfoque para el infinito cuando $p \rightarrow 1$ (ver Figura 2.1).

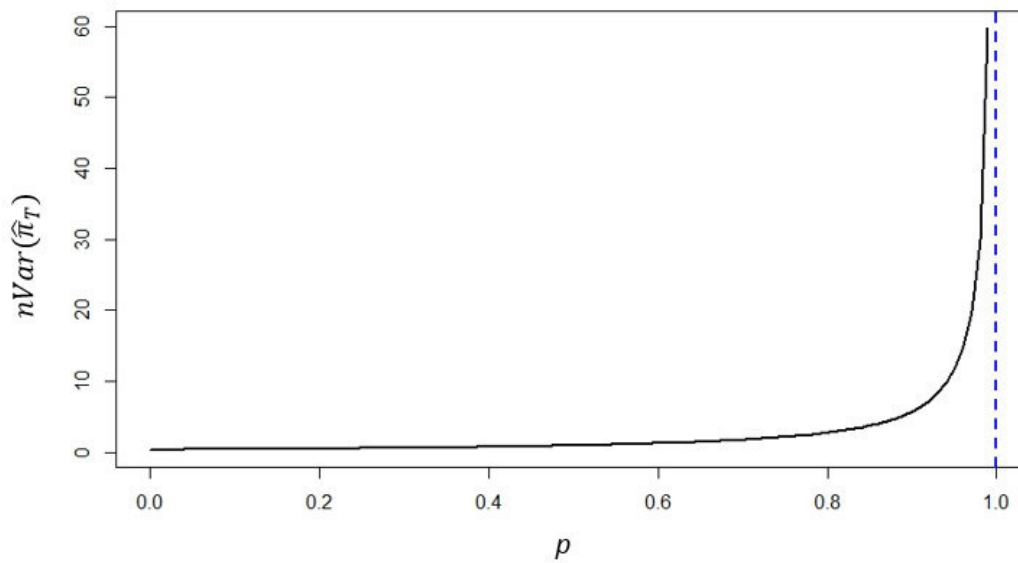


Figura 2.1. Gráfico de $nVar(\hat{\pi}_T)$ definido en 2.5 versus p con $\pi = 0.40$ para el modelo triangular.

2.2.5. Eficiencia relativa

La eficiencia relativa del diseño triangular con respecto al diseño de entrevista directa es:

$$ER_{T \rightarrow D}(\pi, p) = \frac{Var(\hat{\pi}_T)}{Var(\hat{\pi}_D)} = 1 + \frac{p}{(1-p)\pi}, \quad (2.7)$$

Que es independiente del tamaño de la muestra n . La Tabla 2.2 muestra algunos valores de la eficiencia relativa, $ER_{T \rightarrow D}(\pi, p)$, para diversas combinaciones de π y p . Por ejemplo, cuando $\pi = 0.40$ y $p = 5/12$, tenemos $ER_{T \rightarrow D}(0.4, 5/12) = 2.7857$, lo que indica que el tamaño de muestra requerido para el diseño triangular es de aproximadamente 2.8 veces

que la requerida para la entrevista directa para conseguir la misma precisión de estimación.

Tabla 2.2. Eficiencia relativa $ER_{T \rightarrow D}(\pi, p)$ para varias combinaciones de π y p .

π	P						
	0.20	0.30	0.34	0.38	0.42	0.46	0.50
0.05	6.000	9.571	11.303	13.258	15.286	18.037	21.000
0.10	3.500	5.286	6.152	7.129	8.143	9.519	11.000
0.15	2.667	3.857	4.434	5.086	5.762	6.679	7.667
0.20	2.250	3.143	3.576	4.065	4.571	5.259	6.000
0.25	2.000	2.714	3.061	3.452	3.857	4.407	5.000
0.30	1.833	2.429	2.717	3.043	3.381	3.840	4.333
0.35	1.714	2.225	2.472	2.751	3.041	3.434	3.857
0.40	1.625	2.071	2.288	2.532	2.786	3.130	3.500
0.45	1.556	1.952	2.145	2.362	2.587	2.893	3.222
0.50	1.500	1.857	2.030	2.226	2.429	2.704	3.000

2.2.6. Grado de protección de la privacidad del entrevistado

Investigando el grado de protección de la privacidad (GPP). Intuitivamente, el grado óptimo de protección de la privacidad del entrevistado se alcanza con $p \approx 0.5$, cuando π es muy pequeño. Dado que la privacidad de la información divulgada por el entrevistado con respecto a su pertenencia a la clase sensible $\{Y = 1\}$ se caracteriza a través $P(Y = 1 | Y_T = 0)$ y $P(Y = 1 | Y_T = 1)$, tenemos:

$$GPP_O(\pi, p) \triangleq P(Y = 1 | Y_T = 0) = 0 \quad (2.8)$$

$$GPP_{\blacksquare}(\pi, p) \triangleq P(Y = 1 | Y_T = 1) = \frac{\pi}{\pi + (1 - \pi)p} \quad (2.9)$$

En particular, cuando $p = 0$ tenemos que el $GPP_{\blacksquare}(\pi, 0) = 1$, que corresponde al diseño de entrevista directa. Cuando $p = 1$, se obtiene $GPP_{\blacksquare}(\pi, 1) = \pi$, que se corresponde con el caso de que $W = 1$; es decir, el entrevistado tiene un cumpleaños entre agosto y diciembre.

Cuando $p = 0.5$.

$$GPP_{\blacksquare}(\pi, 0.5) = \frac{2\pi}{\pi + 1} > \pi$$

La figura 2.2 muestra que para cualquier π fijo, $GPP_{\blacksquare}(\pi, p)$ es una función monótona decreciente de p con máxima 1 y el mínimo π .

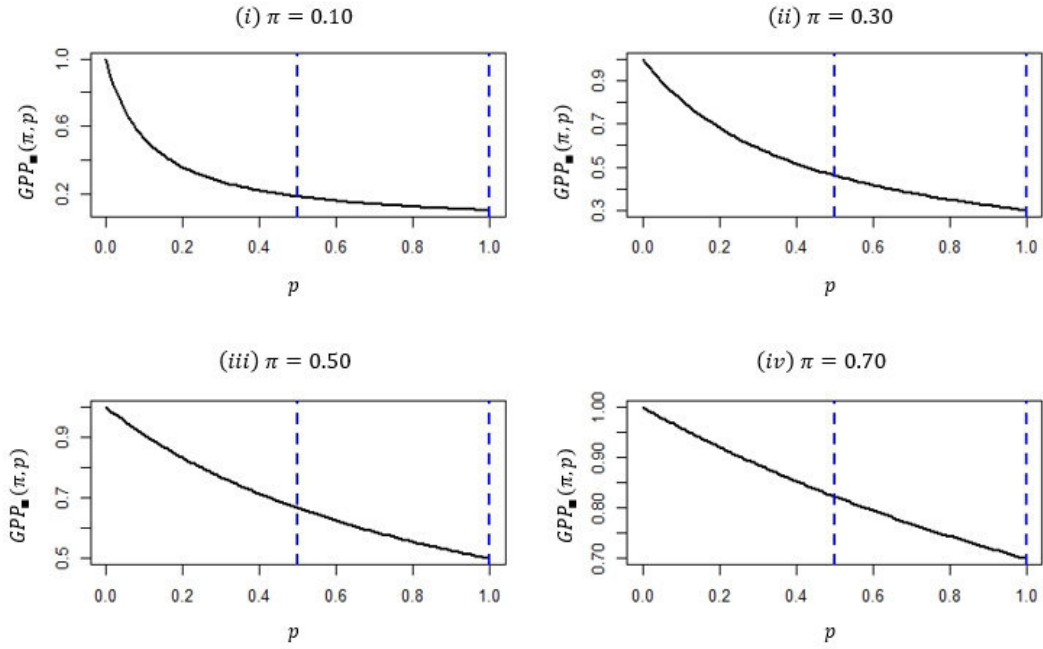


Figura 2.2. Gráfico de $GPP_{\blacksquare}(\pi, p)$ definido en 2.9 versus p para el modelo triangular para un π fijo. (i) $\pi = 0.10$; (ii) $\pi = 0.30$; (iii) $\pi = 0.50$; (iv) $\pi = 0.70$.

2.3. COMPARACIÓN DEL MODELO TRIANGULAR CON EL MODELO DE RESPUESTA ALEATORIZADA DE WARNER.

2.3.1. La diferencia de dos varianzas.

La equivalencia entre el modelo de Warner y el modelo de entrevista directa (Tian & Tang, 2014) proporciona una base para comparar el modelo de Warner con el modelo triangular, no aleatorizado.

A partir de:

$$Var(\hat{\pi}_W) = Var(\hat{\pi}_D) + \frac{p(1-p)}{n(2p-1)^2} \quad (2.10)$$

$$Var(\hat{\pi}_D) = \frac{\pi(1-\pi)}{n}$$

Utilizando el criterio de varianza, se obtiene:

$$Var(\hat{\pi}_W) - Var(\hat{\pi}_T) = \frac{p}{n(1-p)(2p-1)^2} * h_{WT}(p|\pi) \quad (2.11)$$

Donde:

$$h_{WT}(p|\pi) = (4\pi - 3)p^2 + (2 - 4\pi)p + \pi$$

Es una función cuadrática de p para un π fijo ($\pi \neq 3/4$). Tenemos el siguiente resultado.

Teorema 2.1. El modelo triangular siempre es más eficiente que el modelo de Warner para cualquier $\pi \in (0, 1)$ y $p < 2/3$. Específicamente,

(1) Si $\pi = 3/4$, entonces:

$$Var(\hat{\pi}_W) \geq Var(\hat{\pi}_T), \text{ para } 0 < p \leq 3/4 \text{ (} p \neq 0.5 \text{)} \quad (2.12)$$

(2) Si $\pi \neq 3/4$, entonces:

$$Var(\hat{\pi}_W) \geq Var(\hat{\pi}_T), \text{ para } 0 < p \leq p_\pi \text{ (} p \neq 0.5 \text{)} \quad (2.13)$$

Donde.

$$p_\pi = \frac{2\pi - 1 - \sqrt{1 - \pi}}{4\pi - 3}$$

Es una función creciente de π y alcanza su mínimo de $2/3$ a $\pi = 0$.

Prueba.

(1) Si $\pi = 3/4$, entonces $h_{WT}(p|\pi) = 3/4 - p$ no es negativo cuando $0 < p \leq 3/4$.

De 2.11, obtenemos 2.12, inmediatamente.

(2) Caso I: $\pi > 3/4$, tenemos. $h_{WT}(p|\pi) \geq 0$ para $p \leq p_\pi$ o $p \geq p_2$

Donde:

$$p_2 \triangleq \frac{2\pi - 1 + \sqrt{1 - \pi}}{4\pi - 3}$$

Note que:

$$\frac{dp_\pi}{d\pi} = \frac{(\sqrt{1-\pi} - 2)^2}{2\sqrt{1-\pi} (4\pi - 3)^2} > 0$$

Por lo tanto, p_π es una función creciente de π .

Es fácil demostrar que $0.5 < p_\pi < 1 < p_2$. (2.13) sigue inmediatamente.

Caso II: $\pi < 3/4$. Del mismo modo, tenemos:

$$h_{WT}(p|\pi) \geq 0 \text{ para } p_1 \leq p \leq p_\pi,$$

Donde:

$$p_1 = \frac{-(2\pi - 1 + \sqrt{1-\pi})}{3 - 4\pi}$$

Ahora tenemos que:

$$p_1 < 0 < \frac{1}{2} < p_\pi < 1.$$

Por lo tanto sigue (2.13).

2.3.2. Eficiencia relativa del modelo de Warner y el modelo triangular

En este sentido, consideramos que la eficiencia relativa (ER) del modelo Warner ($p \neq 0.5$) para el modelo triangular; es decir,

$$ER_{w \rightarrow T}(\pi, p) = \frac{Var(\hat{\pi}_W)}{Var(\hat{\pi}_T)} = \frac{\pi + p(1-p)/\{(2p-1)^2(1-\pi)\}}{\pi + p/(1-p)}$$

La ER es independiente del tamaño de la muestra n y sólo depende de los parámetros π y p . Observamos que los entrevistadores suelen seleccionar p en $[0.20, 0.46]$ para el modelo de Warner. De la Tabla 2.3, cuando $0.2 \leq p \leq 0.46$ (o $0.54 \leq p \leq 0.66$), la eficiencia de la estrategia triangular es de aproximadamente 1.7 hasta 222 (o 1.2 hasta 187) veces mayor que la del modelo de Warner. En particular, cuando $0.49 \leq p \leq 0.51$ (que son el rango óptimo para que la privacidad de los encuestados esté protegida), la eficiencia de la estrategia triangular es de aproximadamente 600 hasta 3358 veces mayor que la del modelo de Warner.

Tabla 2.3. Eficiencia relativa $ER_{W \rightarrow T}(\pi, p)$ para varias combinaciones de π y p .

π	P													
	0.2	0.3	0.34	0.38	0.46	0.49	0.51	0.54	0.56	0.58	0.583	0.6	0.63	0.66
0.05	1.72	2.99	4.17	6.57	45.35	650.6	602.9	33.4	13.6	7.03	6.38	4.1	2.09	1.18
0.1	1.69	2.94	4.12	6.51	45.41	654.4	608.5	33.9	13.9	7.21	6.55	4.23	2.18	1.24
0.2	1.67	2.92	4.11	6.53	46.31	672.9	629.5	35.4	14.6	7.65	6.96	4.52	2.37	1.37
0.3	1.69	2.98	4.2	6.72	48.39	708.1	665.8	37.8	15.7	8.27	7.53	4.93	2.61	1.53
0.4	1.75	3.12	4.42	7.12	51.99	765.4	722.9	41.3	17.2	9.13	8.32	5.47	2.92	1.73
0.5	1.85	3.36	4.8	7.79	57.79	855.7	811.2	46.6	19.5	10.3	9.47	6.25	3.35	2
0.6	2.01	3.77	5.45	8.92	67.24	1001	952.2	55	23.1	12.3	11.2	7.43	4	2.39
0.7	2.29	4.49	6.58	10.9	83.81	1254	1196	69.4	29.3	15.5	14.2	9.41	5.07	3.03
0.8	2.87	5.99	8.93	15	117.9	1774	1697	98.7	41.6	22.1	20.2	13.3	7.21	4.29
0.9	4.64	10.5	16.1	27.6	222	3358	3219	187.6	79.1	42.1	38.4	25.3	13.5	8.03

2.3.3. Grado de protección de la privacidad.

Para comparar el GPP (Grado de Protección de la Privacidad) del modelo Warner con el modelo triangular, consideraremos tres casos (es decir $p = 0.3, 0.35$ y 0.4) que son algunas opciones prácticas para el modelo de Warner, y dos casos (es decir, $p = 0.5$ y $7/12$), que son dos opciones óptimas para el modelo triangular. La Tabla 2.4 da GPPs para diversas combinaciones de π y p . Cuando se compara el modelo de Warner con $p = 0.35$ con el modelo triangular con $p = 7/12$, tenemos:

$$GPP_{si} > GPP_o$$

$$GPP_{no} > GPP_{\blacksquare}$$

Para todos π en la tabla. Por lo tanto, el modelo triangular con $p = 7/12$ tiene mayores GPP que el modelo Warner con $p = 0.35$. Podemos llegar a una conclusión similar al comparar el modelo triangular con $p = 0.5$ con el modelo de Warner con $p = 0.30$.

Donde A indica que posee la característica sensible.

$$GPP_{si}(\pi, p) \cong P(A|si) = \frac{\pi p}{\pi p + (1 - \pi)(1 - p)} \quad (2.14)$$

$$GPP_{no}(\pi, p) \cong P(A|no) = \frac{\pi(1 - p)}{\pi(1 - p) + (1 - \pi)p} \quad (2.15)$$

Tabla 2.4. Comparación de GPP para varias combinaciones de π y p .

π	Modelo de Warner						Modelo Triangular	
	p= 0.30		p= 0.35		p= 0.40		p= 0.50	p= 7/12
	GPP_{si}	GPP_{no}	GPP_{si}	GPP_{no}	GPP_{si}	GPP_{no}	GPP_{\blacksquare}	GPP_{\blacksquare}
0.05	0.022	0.109	0.027	0.089	0.033	0.073	0.095	0.082
0.10	0.045	0.205	0.056	0.171	0.068	0.142	0.181	0.160
0.20	0.096	0.368	0.118	0.317	0.142	0.272	0.333	0.300
0.30	0.155	0.500	0.187	0.443	0.222	0.391	0.461	0.423
0.40	0.222	0.608	0.264	0.553	0.307	0.500	0.571	0.533
0.50	0.300	0.700	0.350	0.650	0.400	0.600	0.666	0.631
0.60	0.391	0.778	0.447	0.735	0.500	0.692	0.750	0.720
0.70	0.500	0.845	0.557	0.813	0.609	0.778	0.823	0.800
0.80	0.632	0.903	0.683	0.881	0.727	0.857	0.889	0.873
0.90	0.794	0.955	0.829	0.944	0.857	0.931	0.947	0.939

Nota: GPP_{si} , GPP_{no} , GPP_o y GPP_{\blacksquare} se definen por (2.14), (2.15), (2.8) y (2.9), respectivamente. Además, $GPP_o = 0$ para todo π y p .

2.4. PROPIEDADES ASINTÓTICAS DEL ESTIMADOR DE MÁXIMA VEROSIMILITUD.

2.4.1. Una derivación alternativa del EMV

Supongamos que de n encuestados, los resultados indican que s encuestados marcaron con una cruz en el cuadrado superior y $n - s$ marcan el círculo. (Ver Tabla 2.1).

Mediante la introducción de un nuevo parámetro $\theta = \pi + (1 - \pi)p$, tenemos:

$$\pi = \frac{\theta - p}{1 - p} \quad (2.16)$$

La función de verosimilitud es proporcional a $\theta^s(1 - \theta)^{n-s}$ y el EMV de θ está dada por $\hat{\theta} = s/n$. Por lo tanto, el EMV de π es:

$$\hat{\pi}_T = \frac{\hat{\theta} - p}{1 - p} \quad (2.17)$$

Siempre que $\hat{\pi}_T \in [0,1]$. Cuando $\hat{\pi}_T < 0$ o $\hat{\pi}_T > 1$, podemos utilizar el algoritmo EM

$$\tilde{\pi}_T = \frac{a + Z - 1}{a + b + n - 2} \quad (2.18)$$

$$E(Z|Y_{obs}, \pi) = \frac{s\pi}{\pi + (1 - \pi)p} \quad (2.19)$$

Con $a = b = 1$ (es decir, si la distribución uniforme en $[0,1]$ se adopta como valor apriori de π , el modo posterior de π es igual al EMV de π) para hallar el EMV de π . Note que (2.17) es idéntico al $\hat{\pi}_T$ dado por (2.3). Desde $s \sim \text{Binomial}(n, \theta)$, tenemos:

$$E(s) = n\theta$$

$$\text{Var}(s) = n\theta(1 - \theta)$$

Por lo tanto, $E(\hat{\pi}_T) = \pi$; que es, $\hat{\pi}_T$ insesgado, y

$$\text{Var}(\hat{\pi}_T) = \frac{\theta(1 - \theta)}{n(1 - p)^2} = \frac{\pi(1 - \pi)}{n} + \frac{p(1 - \pi)}{n(1 - p)} \quad (2.20)$$

Es de destacar que la varianza de $\hat{\pi}_T$ se puede expresar como la suma de la varianza debido al muestreo y la varianza debido a la introducción de la variable W no- sensible.

2.4.2. Intervalo de confianza de Wald

Es fácil demostrar que una estimación insesgada de $\text{Var}(\hat{\pi}_T)$ viene dada por

$$\widehat{\text{Var}}(\hat{\pi}_T) = \frac{\hat{\theta}(1 - \hat{\theta})}{n(1 - p)^2} \quad (2.21)$$

Cuando $n \rightarrow \infty$, el teorema de límite central implica que $\hat{\pi}_T$ es asintóticamente normal; eso es,

$$(\hat{\pi}_T - \pi) / \sqrt{\widehat{\text{Var}}(\hat{\pi}_T)} \sim N(0,1) \quad (2.22)$$

De (2.22) la prueba de hipótesis común sobre π se puede establecer fácilmente. El $(1 - \alpha)100\%$ del intervalo de confianza Wald de π está dado por:

$$[\hat{\pi}_{T,WL}, \hat{\pi}_{T,WU}] = \left[\hat{\pi}_T - Z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\pi}_T)}, \hat{\pi}_T + Z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\pi}_T)} \right] \quad (2.23)$$

2.5. EL MODELO TRIANGULAR PARA EL PROBLEMA DE UNA MUESTRA

Para el modelo triangular descrito en la Tabla 2.1, el estimador de máxima verosimilitud de π y su varianza son dados por (2.3), que puede ser reescrito como:

$$\hat{\pi}_T = \frac{\bar{y}_T - p}{1 - p}$$

$$Var(\hat{\pi}_T) = \frac{\theta(1 - \theta)}{n(1 - p)^2} \quad (2.24)$$

Donde

$$\bar{y}_T = \frac{1}{n} \sum_{i=1}^n y_{i,T}$$

$$\theta = \pi + (1 - \pi)p$$

Cuando $n \rightarrow \infty$, el teorema del límite central implica que $\hat{\pi}_T$ es asintóticamente una distribución normal; esto es:

$$\frac{\hat{\pi}_T - \pi}{\sqrt{Var(\hat{\pi}_T)}} = \frac{n\hat{\pi}_T - n\pi}{\sqrt{n\theta(1 - \theta)/(1 - p)}} \sim N(0,1) \quad (2.25)$$

2.5.1. Prueba de hipótesis unilateral

Para probar si hay una diferencia entre la proporción de la población (π) con la característica sensible y un valor de referencia predeterminado (π_0), consideramos las siguientes hipótesis:

$$H_0: \pi = \pi_0 \quad \text{versus} \quad H_1: \pi < \pi_0 \quad (2.26)$$

Bajo la hipótesis nula H_0 , a partir de (2.25), tenemos:

$$\frac{n_T \hat{\pi}_T - n_T \pi_0}{\sqrt{n_T \theta_0 (1 - \theta_0) / (1 - p)}} \sim N(0,1), \quad n_T \rightarrow \infty$$

Donde n_T indica el tamaño de muestra necesario para el modelo triangular para la prueba unilateral, y

$$\theta_0 = \pi_0 + (1 - \pi_0)p$$

Por lo tanto, se rechaza la hipótesis nula H_0 a un nivel de significancia α cuando observamos el siguiente evento.

$$\mathbb{E}_T = \left\{ n_T \hat{\pi}_T \leq n_T \pi_0 - \frac{z_\alpha \sqrt{n_T \theta_0 (1 - \theta_0)}}{1 - p} \right\} \quad (2.27)$$

Donde z_α es el cuantil superior α de la distribución normal estándar. Ahora, suponemos que H_1 es verdadera y $\pi = \pi_1$ con $\pi_1 < \pi_0$. Entonces, el poder de la prueba unilateral es aproximadamente dado por:

$$\begin{aligned} \text{Poder}(\pi_1) &= P(\text{Rechazar } H_0 | \pi = \pi_1) \\ &= P \left\{ \frac{n_T \hat{\pi}_T - E_{H_1}(n_T \hat{\pi}_T)}{\sqrt{\text{Var}_{H_1}(n_T \hat{\pi}_T)}} \leq \frac{n_T \hat{\pi}_0 - z_\alpha \sqrt{n_T \theta_0 (1 - \theta_0)} / (1 - p) - n_T \pi_1}{\sqrt{n_T \theta_1 (1 - \theta_1)} / (1 - p)} \right\} \\ &\approx \Phi \left(\frac{\sqrt{n_T}(\pi_0 - \pi_1)(1 - p) - z_\alpha \sqrt{\theta_0 (1 - \theta_0)}}{\sqrt{\theta_1 (1 - \theta_1)}} \right) \end{aligned} \quad (2.28)$$

Donde

$$\theta_1 = \pi_1 + (1 - \pi_1)p$$

Y $\Phi(\cdot)$ denota la función de distribución acumulada de la distribución normal estándar. Como resultado, el tamaño de muestra necesario para conseguir u obtener una potencia de $1 - \beta$ puede obtenerse resolviendo la siguiente ecuación

$$\sqrt{n_T}(\pi_0 - \pi_1)(1 - p) - z_\alpha \sqrt{\theta_0 (1 - \theta_0)} = z_\beta \sqrt{\theta_1 (1 - \theta_1)}$$

Lo que lleva a

$$n_T = \left\{ \frac{z_\alpha \sqrt{\theta_0 (1 - \theta_0)} + z_\beta \sqrt{\theta_1 (1 - \theta_1)}}{(\pi_0 - \pi_1)(1 - p)} \right\}^2 \quad (2.29)$$

2.5.2. Prueba de hipótesis bilateral

Ahora tenemos en cuenta una prueba de dos colas para las siguientes hipótesis;

$$H_0: \pi = \pi_0 \quad \text{versus} \quad H_1: \pi \neq \pi_0$$

Con la asignación de $\alpha/2$ a cada cola de la región de rechazo. Las relaciones entre el poder, tamaño de la muestra y el tamaño del efecto es aproximadamente dado por:

$$Poder(\pi_1) \approx \Phi \left(\frac{z_{\frac{\alpha}{2}} \sqrt{\theta_o(1-\theta_o)} + z_{\beta} \sqrt{\theta_1(1-\theta_1)}}{(\pi_0 - \pi_1)(1-p)} \right)^2$$

La fórmula tamaño de la muestra correspondiente es:

$$n_{T,2} = \left\lceil \frac{z_{\alpha/2} \sqrt{\theta_o(1-\theta_o)} + z_{\beta} \sqrt{\theta_1(1-\theta_1)}}{(\pi_0 - \pi_1)(1-p)} \right\rceil^2 \quad (2.30)$$

2.5.3. Evaluación del desempeño comparando la potencia exacta y el poder asintótico.

Para examinar la exactitud de la fórmula de la potencia asintótica (2.28), trazamos en la Figura 2.3 la potencia exacta, así como los poderes asintóticos versus el tamaño de la muestra n_T en diversas combinaciones de (π_0, π_1) en $p = 0.5$ y el nivel de significación $\alpha = 0.05$. La potencia exacta (en π_1) para cualquier tamaño de muestra particular n_T es:

$$\begin{aligned} Potencia\ Exacta(\pi_1) &= \sum_{x \in \mathbb{E}_T} Binomial(x | n_T, \pi_1 + (1 - \pi_1)p) \\ &= \sum_{x \in \mathbb{E}_T} \binom{n_T}{x} \theta_1^x (1 - \theta_1)^{n_T - x} \end{aligned} \quad (2.31)$$

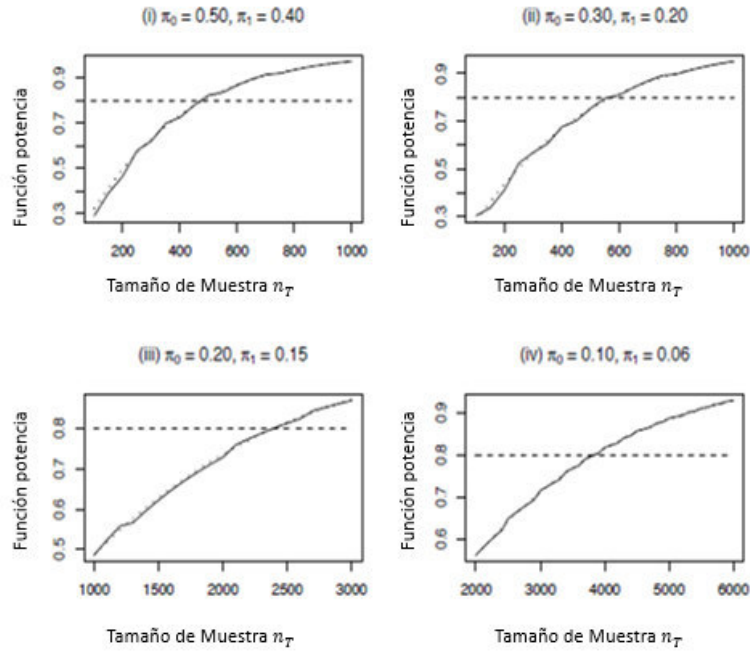


Figura 2.3. Las comparaciones de la potencia exacta (2.31) (indicado por la línea continua) con la potencia asintótica (2.28) (denotados por la línea de puntos) Versus el tamaño de la muestra n_T bajo diversas combinaciones de (π_0, π_1) en $p = 0.5$ y $\alpha = 5\%$. (i) $(\pi_0, \pi_1) = (0.50, 0.40)$; (ii) $(\pi_0, \pi_1) = (0.30, 0.20)$; (iii) $(\pi_0, \pi_1) = (0.20, 0.15)$; (iv) $(\pi_0, \pi_1) = (0.10, 0.06)$.

Donde la región de rechazo \mathbb{E}_T especificada en (2.27) se puede reescribir como:

$$\mathbb{E}_T = \left\{ X_T \leq n_T \theta_o - z_\alpha \sqrt{n_T \theta_o (1 - \theta_o)} \right\}$$

Donde:

$$X_T = \sum_{i=1}^{n_T} y_{i,T} \sim \text{Binomial}(n_T, \theta).$$

En general, encontramos que la función de potencia asintótica (2.25) proporciona una aproximación satisfactoria de la potencia exacta, sobre todo cuando n_T aumenta.

2.5.4. Evaluación del desempeño calculando n_T con n_T/n_D .

La Tabla 2.5 reporta algunos valores de tamaño de la muestra n_T dada por (2.28) con $\alpha = 5\%$ y 80% de potencia en varias combinaciones de (π_0, π_1, p) y la correspondiente potencia exacta calculada de acuerdo a (2.31).

Tabla 2.5. Tamaño de muestra n_T para prueba de hipótesis $H_0: \pi = \pi_0$ versus $H_1: \pi = \pi_1 < \pi_0$ con $\alpha = 5\%$, 80 % de potencia y razón n_T/n_D .

π_0	π_1	p=0.00	p=0.42		p=0.50		p=0.58	
		n_D	n_T	n_T/n_D	n_T	n_T/n_D	n_T	n_T/n_D
0.5	0.4	153	389	2.54	483	3.15	614	4.01
	0.35	67	175	2.61	218	3.25	278	4.14
	0.3	37	100	2.7	125	3.37	159	4.29
0.4	0.35	583	1673	2.86	2109	3.61	2719	4.66
	0.3	142	422	2.97	534	3.76	691	4.86
	0.25	61	189	3.09	240	3.93	312	5.11
0.3	0.25	501	1767	3.52	2274	4.53	2983	5.95
	0.2	119	444	3.73	573	4.81	755	6.34
	0.18	81	309	3.81	399	4.92	527	6.5
0.2	0.16	584	2831	4.84	3729	6.38	4986	8.53
	0.13	181	925	5.11	1222	6.75	1637	9.04
	0.1	83	453	5.45	600	7.22	806	9.71
0.1	0.08	1303	11317	8.68	15321	11.7	20927	16.1
	0.06	301	2826	9.38	3834	12.7	5246	17.4
	0.04	121	1254	10.3	1706	14.1	2338	19.3

Fuente: Table 2 de Tian et al. (2011).

Nota: n_D es el tamaño de muestra para el diseño de entrevista directa, dado en 2.32.

Generalmente el tamaño de muestra n_T es suficiente para garantizar la potencia deseada de la prueba; que es:

$$|Potencia actual - potencia deseada| \leq 5\%$$

Para un par dado de (π_0, π_1) , podemos ver que n_T es una función creciente de p . En particular, cuando $p = 0$, el diseño triangular se reduce al diseño de entrevista directa. Aquí, se denota el tamaño de la muestra del diseño de entrevista directa entrevistado por n_D . En (2.26), el estableciendo $p = 0$, obtenemos

$$n_D = \left\{ \frac{z_\alpha \sqrt{\pi_0(1 - \pi_0)} + z_\beta \sqrt{\pi_1(1 - \pi_1)}}{\pi_0 - \pi_1} \right\}^2 \quad (2.32)$$

En la Tabla 2.5, también presenta las proporciones de n_T/n_D para $p = 0.42, 0.5, 0.58$, respectivamente. Por ejemplo, cuando $(\pi_0, \pi_1) = (0.4, 0.3)$ y $p = 0.42$, tenemos $n_T/n_D = 2.97$, lo que indica que el tamaño de muestra necesario para el diseño triangular es

de aproximadamente 3 veces que la requerida para el diseño de entrevista directa a fin de lograr la misma potencia para la prueba unilateral. Aquí, elegimos la variable dicotómica W no sensible como la duración de cumpleaños del entrevistado. Para $p = 0.42$ (es decir, $5/12$), 0.50 (es decir, $6/12$) y 0.58 (es decir, $12/07$), que representan $W = 1$ si el entrevistado nació entre enero y mayo, enero y junio, y enero y julio, respectivamente; y $W = 0$ en caso contrario.

2.6. EL MODELO TRIANGULAR PARA EL PROBLEMA DE DOS MUESTRAS.

Cuando llevamos a cabo dos estudios independientes sobre una cuestión sensible común en dos poblaciones diferentes / regiones (indexados por $k = 1, 2$) con el diseño triangular presentado es de interés para decidir el diseño, el número de encuestados en dos grupos con el fin de comparar las proporciones ($\pi_k, k = 1, 2$) de los sujetos con un atributo sensible. Para un k fijo, definimos una variable oculta $Y_{k,T}$ como sigue:

$$Y_{k,T} = \begin{cases} 1, & \text{si la cruz está en el cuadrado superior de la } k - \text{ésima región.} \\ 0, & \text{si la cruz está en el círculo de la } k - \text{ésima región.} \end{cases}$$

Donde $p_k = P(W_k = 1), k = 1, 2$ se supone que son conocidas pero no necesariamente iguales. Tenemos:

$$P(Y_{k,T} = 1) = \pi_k + (1 - \pi_k) p_k \text{ y } P(Y_{k,T} = 0) = (1 - \pi_k)(1 - p_k).$$

Sea $Y_{obs} = \{y_{ik,T} : i = 1, \dots, n_k, k = 1, 2\}$ denotando los datos observados para los $n_1 + n_2$ encuestados en las dos regiones, donde $y_{ik,T} = 1$ (o *cero*) si el encuestado i en la k -ésima región pone una cruz en el cuadrado superior (o en el círculo). Supongamos que $\rho = n_1/n_2$. La función de probabilidad para π_1 y π_2 entonces dada por:

$$L_T(\pi_1, \pi_2 | Y_{obs}) = \prod_{k=1}^2 \prod_{i=1}^{n_k} \{\pi_k + (1 - \pi_k) p_k\}^{y_{ik,T}} \{(1 - \pi_k)(1 - p_k)\}^{1-y_{ik,T}}$$

El EMV resultante de π_k y la varianza correspondientes están dadas por:

$$\hat{\pi}_k = \frac{\bar{y}_{k,T} - p_k}{1 - p_k} \text{ y } Var(\hat{\pi}_k) = \frac{\vartheta_k(1 - \vartheta_k)}{n_k(1 - p_k)^2} \quad k = 1, 2.$$

Donde:

$$\bar{y}_{k,T} = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} y_{ik,T}$$

$$\vartheta_k = \pi_k + (1 - \pi_k)p_k$$

Por tanto

$$\widehat{Var}(\hat{\pi}_k) = \frac{\bar{y}_{k,T}(1-\bar{y}_{k,T})}{n_k (1-p_k)^2}$$

Es el estimador de máxima verosimilitud de $Var(\hat{\pi}_k)$.

Ahora, tenemos en cuenta la siguiente hipótesis bilateral o de dos colas.

$$H_0: \pi_1 = \pi_2 \text{ versus } H_0: \pi_1 \neq \pi_2$$

La hipótesis nula es rechazada con un nivel de significancia α , si

$$\left| \frac{\hat{\pi}_1 - \hat{\pi}_2}{\widehat{SE}_+} \right| > z_{\alpha/2}$$

Donde

$$\widehat{SE}_+ = \sqrt{\sum_{k=1}^2 \widehat{Var}(\hat{\pi}_k)}$$

Denota el estimador de:

$$SE_+ = \sqrt{\sum_{k=1}^2 Var(\hat{\pi}_k)}$$

Bajo la hipótesis alternativa de que $\pi_1 - \pi_2 \neq 0$, la potencia de la prueba anterior es aproximadamente dado por:

$$\Phi \left(\frac{|\pi_1 - \pi_2| - \frac{z_{\alpha}}{2} * \widehat{SE}_+}{SE_+} \right)$$

Siguiendo Chow, Shao y Wang (2003), que se aproximan aún más al poder por:

$$\Phi\left(\frac{|\pi_1 - \pi_2|}{SE_+} - z_{\alpha/2}\right)$$

Por lo tanto, el tamaño de muestra necesario para obtener una potencia deseada de $1 - \beta$ se puede obtener mediante la resolución de la ecuación:

$$\frac{|\pi_1 - \pi_2|}{SE_+} - z_{\frac{\alpha}{2}} = z_{\beta}$$

Esto lleva a:

$$n_1 = \rho n_2$$

Y

$$n_2 = \frac{\left(z_{\frac{\alpha}{2}} + z_{\beta}\right)^2}{(\pi_1 - \pi_2)^2} \left\{ \frac{\vartheta_1(1 - \vartheta_1)}{\rho(1 - p_1)^2} + \frac{\vartheta_2(1 - \vartheta_2)}{\rho(1 - p_2)^2} \right\}$$

CAPÍTULO III: APLICACIÓN DEL MTRNA

3.1. IMPLEMENTACIÓN DEL MODELO TRIANGULAR DE RESPUESTA NO ALEATORIZADA.

El modelo triangular de respuesta no aleatorizada propuesto por (Yu et al., 2008) fue utilizado para realizar una encuesta sobre: Copia en los exámenes, consumo de drogas para mejorar el rendimiento académico, uso de notas en una examen, copia de un párrafo intencionalmente adoptado del trabajo de otra persona.

El propósito de esta investigación es mostrar la utilidad del modelo de respuesta no aleatorizada, para obtener una estimación de algunas variables como (Copia en los exámenes, uso de notas en un examen, consumo de drogas para mejorar el rendimiento académico y copia de párrafo intencional adoptado del trabajo de otra persona). De tal manera se irá reduciendo principalmente el sesgo de respuesta (por respuestas deliberadamente falsas) y se obtendrán tasas de respuestas mayores.

3.1.1. Elección de las variables.

Las variables elegidas para el estudio fueron las siguientes:

- **Copia en los exámenes:**

Preguntas sensibles

“He copiado de otros estudiantes durante un examen (práctica) por lo menos una vez.”
(Y=1)

“Nunca he copiado de otros estudiantes durante un examen (práctica).” (Y=0)

Preguntas no sensibles

“Nací entre Agosto y Diciembre.” (W=1)

“No nací entre Agosto y Diciembre.”(W=0)

- **Uso de notas en un examen:**

Preguntas sensibles

“He utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares).” (Y=1)

“Nunca he utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares).” (Y=0)

Preguntas no sensibles

“He interrumpido mis estudios universitarios por lo menos una vez” (W=1)

“Nunca he interrumpido mis estudios universitarios” (W=0)

• Consumo de drogas para mejorar el rendimiento académico

Preguntas sensibles:

“He utilizado drogas para mejorar mi rendimiento académico en un examen o práctica por lo menos una vez” (Y=1)

“Nunca he utilizado drogas para mejorar mi rendimiento académico en un examen o práctica” (Y=0)

Preguntas no sensibles:

“Ingrese por la Modalidad de Centro Pre Universitario” (W=1)

“No ingrese por la Modalidad de Centro Pre Universitario” (W=0)

• Copia de un párrafo intencionalmente adoptado del trabajo de otra persona.

Preguntas sensibles

“He presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona por lo menos una vez” (Y=1)

“Nunca he presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona” (Y=0)

Preguntas no sensibles

“Me gusta estudiar solo(a)” (W=1)

“No me gusta estudiar solo(a)” ($W=0$)

Cada una de las preguntas mencionadas fueron organizadas en una tabla de contingencia. (Ver anexo – Cuestionario 1)

3.1.2. Tamaño de muestra:

De acuerdo a Registros Académicos de la Dirección Académica los Alumnos Matriculados el Semestre 2015 - I en la Facultad de Ciencias Matemáticas de la UNMSM, está dividida en cuatro Escuelas académico Profesionales (Ver Tabla 3.1), de acuerdo a esta información se decidió considerar a cada Escuela Académico Profesional como un estrato, en total tenemos 4 estratos o Escuelas. El esquema de muestreo que se utilizó fue el Muestreo Aleatorio Estratificado, con afijación proporcional de acuerdo a la cantidad de alumnos en cada uno de los estratos, donde cada estrato es una Escuela Académico Profesional de la Facultad de Ciencias Matemáticas de la UNMSM.

Tabla 3.1: Distribución de estudiantes de Pre- Grado según Escuela Académico Profesional en el semestre 2015-I-FCM – UNMSM.

Escuela Académico Profesional	Frecuencia	Porcentaje
Matemática	458	30.11%
Estadística	291	19.13%
Investigación Operativa	471	30.97%
Computación Científica	301	19.79%
Total	1521	100.00%

Fuente: Dirección Académica – FCM – UNMSM.

Se utilizó el Muestreo Aleatorio Estratificado (Cochran, W., 1977) con un nivel de confianza del 95% y una potencia de 80 %, el tamaño de muestra fue de 300 alumnos matriculados el Semestre 2015-I. El Modelo de Respuesta No Aleatorizada requiere un tamaño de muestra más grande que el método convencional. La distribución de la muestra por Escuela Académico Profesional se muestra en el Tabla 3.2.

Tabla 3.2: Distribución de la muestra de estudiantes de Pre- Grado según Escuela Académico Profesional en el Semestre 2015-I-FCM – UNMSM.

Escuela Académico Profesional	Frecuencia
Matemática	90
Estadística	58
Investigación Operativa	93
Computación Científica	59
Total	300

Elaboración propia.

El cálculo del tamaño de la muestra es un componente esencial en la encuesta diseñada en temas sensibles. En este capítulo, vamos a discutir el cálculo del tamaño de la muestra para el modelo triangular basado en el método de análisis de potencia para una muestra y dos muestras. En esta sección, consideramos una muestra y se derivan el tamaño de la muestra con fórmulas para el modelo triangular, tanto para pruebas de una y dos colas sobre la base de la muestra grande con aproximación normal.

Tabla 3.3: Probabilidad de presentar la característica no sensible.

Preguntas No Sensibles	Proporción
“Nacidos entre Agosto y Diciembre.”	5/12 =0.417
“He interrumpido mis estudios universitarios por lo menos una vez”.	0.19
“Ingrese por la Modalidad de Centro Pre Universitario”	0.23
“Me gusta estudiar solo(a)”	0.22

Fuente: PERÚ, II Censo Nacional Universitario 2010.

Se utilizó el muestreo aleatorio estratificado con un nivel de confianza del 95 % y una potencia de 80 %. Determinando el tamaño de muestra establecido, se establece considerando

$$\pi_0 = 0.47 \text{ y } \pi_1 = 0.35$$

Hallando el tamaño de muestra de acuerdo a las probabilidades de las preguntas no sensibles se tiene:

Tabla 3.4: Tamaño de muestra de acuerdo a la pregunta no sensible.

Preguntas No Sensibles	Muestra
“Nacidos entre Agosto y Diciembre.”	300
“He interrumpido mis estudios universitarios por lo menos una vez”.	173
“Ingresa por la Modalidad de Centro Pre Universitario”	190
“Me gusta estudiar solo(a)”	186

De la Tabla 3.4 se escoge al tamaño de muestra que es 300. Que fue obtenido de la siguiente manera

Hallando el tamaño de muestra:

$$H_0: \pi = 0.47 \quad \text{versus} \quad H_1: \pi = 0.35 < 0.47$$

Con un $\alpha = 0.05$ y $1 - \beta = 0.80$.

$$n_T = \left\{ \frac{z_\alpha \sqrt{\theta_0(1 - \theta_0)} + z_\beta \sqrt{\theta_1(1 - \theta_1)}}{(\pi_0 - \pi_1)(1 - p)} \right\}^2$$

Donde:

$$\pi_0 = 0.47 \quad \text{y} \quad \pi_1 = 0.35$$

$$p = 5/12 = 0.417$$

$$z_\alpha = 1.645$$

$$z_\beta = 0.842$$

$$\theta_0 = \pi_0 + (1 - \pi_0)p \quad \text{y} \quad \theta_1 = \pi_1 + (1 - \pi_1)p$$

Por lo tanto,

$$n_T = \left\{ \frac{1.645 \sqrt{0.689(1 - 0.689)} + 0.842 \sqrt{0.621(1 - 0.621)}}{(0.47 - 0.35)(1 - 0.417)} \right\}^2$$

$$n_T = 299.01 \approx 300$$

3.1.3. Plan de muestreo de la aplicación.

a) Población objetivo:

La población objetivo, es aquella acerca de la cual deseamos recolectar la información, para nuestro caso, es la población de alumnos matriculados el Semestre 2015-I, en la Facultad de Ciencias Matemáticas, de la UNMSM.

b) Marco Muestral:

El Marco muestral está constituido por el Listado de Alumnos Matriculados el Semestre Académico 2015-I, en la Facultad de Ciencias Matemáticas de la UNMSM, esta base de datos fue proporcionada por Dirección Académica de dicha Facultad.

c) Diseño de la Muestra:

Se dividió el marco muestral en 4 estratos:

Estratos	Escuela Académico Profesional (EAP)
1	Matemática
2	Estadística
3	Investigación Operativa
4	Computación Científica

Para elegir las unidades en el estrato 1, se enumeró el listado de la EAP de Matemática desde 1 hasta 458. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 21:

Datos / Seleccionar Casos / Muestra Aleatoria – muestra / Exactamente 90 de 483 / continuar / aceptar.

De manera similar, en el estrato 2, se enumeró el listado de la EAP de Estadística desde 1 hasta 291. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 21:

Datos / Seleccionar Casos / Muestra Aleatoria – muestra / Exactamente 58 de 291 /continuar / aceptar.

En el estrato 3, se enumeró el listado de la EAP de Investigación Operativa desde 1 hasta 471. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 21:

Datos / Seleccionar Casos / Muestra Aleatoria – muestra Exactamente 93 de 471 / continuar / aceptar.

En el estrato 4, se enumeró el listado de la EAP de Computación Científica desde 1 hasta 301. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 21:

Datos / Seleccionar Casos / Muestra Aleatoria – muestra / Exactamente 59 de 301 / continuar / aceptar.

El esquema de muestreo empleado en cada estrato fue el muestreo aleatorio simple sin reposición.

d) Instrumento de recolección de información:

El cuestionario utilizado se dividió en 2 partes: En una primera parte se registran los datos generales del Alumno seleccionado (Características Generales), los cuales corresponden a su edad, género, EAP y Ciclo que está estudiando. La segunda parte corresponde a las características específicas relacionado al tema de las preguntas sensibles, en donde se registra la respuesta del entrevistado.

El cuestionario utilizado se muestra en el Anexo - Cuestionario 3.

e) Actividades de Campo

Personal de la encuesta y Materiales del trabajo de campo: Para el trabajo de campo se contó con el ambiente de la Oficina 211, cuyos miembros (las profesoras de esa oficina nos concedieron ese ambiente para llevarse a cabo la encuesta). Dos alumnos de Seminario de Tesis, un profesional de psicología y la autora de la tesis, fueron los entrevistadores.

Para la difusión de los alumnos seleccionados, se colocó un Comunicado, con la lista de alumnos elegidos según Escuela Académico Profesional, en una pizarra ubicada en la entrada de la FCM para que ellos pudieran enterarse que habían sido elegidos.

La entrevista se realizó en la Oficina 211, en el horario de 10am a 8pm, la cual empezó el día 17 de junio y finalizó el 3 de julio del año 2015.

Los entrevistadores recibieron charlas, sobre las definiciones utilizadas en el estudio, la explicación del cuestionario y el procedimiento aleatorizado.

Si un alumno seleccionado se presentaba; primeramente debería identificarse con su carnet universitario, el entrevistador debía identificar la Escuela y sus nombres en las listas, las cuales contenían los nombres de los alumnos seleccionados por Escuela y marcando con plumón resaltador, para que quede registrado que ese alumno ya había sido entrevistado. Luego un entrevistador procedía con la explicación del cuestionario, para que este sea correctamente llenado. Cuando el entrevistado terminaba de completar sus respuestas, su cuestionario procedía a ser revisado por el entrevistador para corroborar que no presente errores. Y se le obsequiaba un lapicero y/o chocolate como agradecimiento de su participación.

3.2. RESULTADOS DE LA APLICACIÓN DEL MODELO TRIANGULAR DE RESPUESTA NO ALEATORIZADA.

3.2.1. Evaluación del rendimiento de la muestra.

Se logró entrevistar a 300 alumnos seleccionados, lo cual significó una tasa de respuesta del 100%.

3.2.2. Características de la población investigada

La población investigada es de 1521 alumnos matriculados el Semestre 2015-I, cuyas edades fluctúan entre 16 y 66 años.

3.2.3. Resultados numéricos.

Utilizando las fórmulas de los estimadores de la proporción de respondientes en la población con la pregunta sensible del Modelo Triangular de Respuesta No Aleatorizada y las varianzas del estimador, presentada anteriormente, se obtuvieron los siguientes resultados utilizando el software R-Studio 3.1.2. (Ver Anexo 2)

a) **Copia en los exámenes:**

Sea $\hat{\pi}_T$ el estimador del modelo triangular de respuesta no aleatorizada para la “Proporción de alumnos de la Facultad de Ciencias Matemáticas que han copiado de otros estudiantes durante un examen (práctica) por lo menos una vez.”, cuyo estimador fue desarrollado en el capítulo anterior (ver ecuación 2.3, 2.4 y 2.5):

Luego para el caso en estudio:

Se tiene que: $P[\text{Nací entre Agosto y Diciembre.}] = p_1 = 5/12$,

El promedio de marcar con una cruz el cuadrado superior en la muestra fue:

$$\bar{y}_t = 0.597.$$

Por lo tanto:

$$\hat{\pi}_T = 0.3086$$

De aquí, se obtiene la proporción de que los alumnos que hayan copiado de otros estudiantes durante un examen (práctica) por lo menos una vez es: 0.3086 (30.86%).

Varianza: Se obtiene el siguiente valor:

$$Var(\hat{\pi}_T) = 0.002357442$$

Y la desviación estándar es:

$$\sqrt{Var(\hat{\pi}_T)} = \sqrt{0.002357442} = 0.0485535$$

Hallando los intervalos de confianza:

$$[\hat{\pi}_{T,WL}, \hat{\pi}_{T,WU}] = [0.228736599, \quad 0.388463401]$$

b) **Uso de notas en un examen:**

De igual forma, se tiene que: $P[\text{Haber interrumpido mis estudios universitarios por lo menos una vez.}] = p_2 = 0.19$

El promedio de marcar con una cruz el cuadrado superior en la muestra fue: 0.633

De aquí, se obtiene la proporción de que los alumnos hayan utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares) es: 0.5473 (54.73%).

$$\hat{\pi}_T = 0.5473$$

Varianza

$$Var(\hat{\pi}_T) = 0.001179811$$

Y la desviación estándar es:

$$\sqrt{Var(\hat{\pi}_T)} = \sqrt{0.001179811} = 0.0343484$$

Hallando los intervalos de confianza:

$$[\hat{\pi}_{T,WL}, \hat{\pi}_{T,WU}] = [0.49080191, 0.60379809]$$

c) **Consumo de drogas para mejorar el rendimiento académico**

Se tiene que: P [Ingrese por la Modalidad de Centro Pre Universitario] = $p_3 = 0.23$

El promedio de marcar con una cruz el cuadrado superior en la muestra fue: 0.327

De aquí, se obtiene la proporción de que los alumnos hayan utilizado drogas para mejorar su rendimiento académico en un examen o práctica por lo menos una vez es: 0.1255 (12.55%).

$$\hat{\pi}_T = 0.1255$$

Varianza

$$Var(\hat{\pi}_T) = 0.001236609$$

Y la desviación estándar es:

$$\sqrt{Var(\hat{\pi}_T)} = \sqrt{0.001236609} = 0.03516545$$

Hallando los intervalos de confianza:

$$[\hat{\pi}_{T,WL}, \hat{\pi}_{T,WU}] = [0.067657982, 0.183342018]$$

d) **Copia de un párrafo intencionalmente adoptado del trabajo de otra persona.**

Se tiene que: $P[\text{Me gusta estudiar solo(a)}] = p_3 = 0.22$.

El promedio de marcar con una cruz el cuadrado superior en la muestra fue: 0.78

De aquí, se obtiene la proporción de que los alumnos hayan presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona por lo menos una vez.es: 0.7179 (71.79%).

$$\hat{\pi}_T = 0.7179$$

Varianza

$$Var(\hat{\pi}_T) = 0.0009401709$$

Y la desviación estándar es:

$$\sqrt{Var(\hat{\pi}_T)} = \sqrt{0.0009401709} = 0.030662206$$

Hallando los intervalos de confianza:

$$[\hat{\pi}_{T,WL}, \hat{\pi}_{T,WU}] = [0.667465159, 0.768334841]$$

3.3. IMPLEMENTACIÓN DEL MODELO DE ENTREVISTA DIRECTA.

Para la implementación del modelo de entrevista directa utilizaremos:

$$\pi_0 = 0.47 \text{ y } \pi_1 = 0.35$$

A fin de comparar con el modelo triangular de respuesta no aleatorizada.

Hallando el tamaño de muestra:

$$H_0: \pi = 0.47 \quad \text{versus} \quad H_1: \pi = 0.35 < 0.47$$

Con un $\alpha = 0.05$ y $1 - \beta = 0.80$.

Dado que en la entrevista directa no se considera la pregunta no sensible: $p = 0$

$$n_D = \left\{ \frac{z_\alpha \sqrt{\theta_o(1 - \theta_o)} + z_\beta \sqrt{\theta_1(1 - \theta_1)}}{(\pi_0 - \pi_1)} \right\}^2 = 110.97 \approx 111$$

Donde:

$$\pi_0 = 0.47$$

$$\pi_1 = 0.35$$

$$z_\alpha = 1.645$$

$$z_\beta = 0.842$$

$$\theta_0 = \pi_0 + (1 - \pi_0)p \quad y \quad \theta_1 = \pi_1 + (1 - \pi_1)p$$

3.3.1. Plan de muestreo de la aplicación.

a) Población objetivo:

La población objetivo es la misma que para el modelo triangular.

b) Marco Muestral:

De igual manera, el marco muestral es el mismo que utilizamos para el modelo triangular.

c) Diseño de la Muestra:

Se dividió el marco muestral en 4 estratos:

Estratos	Escuela Académico Profesional (EAP)
1	Matemática
2	Estadística
3	Investigación Operativa
4	Computación Científica

Para elegir las unidades en el estrato 1, se enumeró el listado de la EAP de Matemática desde 1 hasta 458. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 21:

Datos / Seleccionar Casos / Muestra Aleatoria – muestra / Exactamente 34 de 483 / continuar / aceptar.

De manera similar, en el estrato 2, se enumeró el listado de la EAP de Estadística desde 1 hasta 291. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 21:

Datos / Seleccionar Casos / Muestra Aleatoria – muestra / Exactamente 21 de 291 /continuar / aceptar.

En el estrato 3, se enumeró el listado de la EAP de Investigación Operativa desde 1 hasta 471. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 21:

Datos / Seleccionar Casos / Muestra Aleatoria – muestra Exactamente 34 de 471 / continuar / aceptar.

En el estrato 4, se enumeró el listado de la EAP de Computación Científica desde 1 hasta 301. Y utilizando el siguiente procedimiento en el Programa Estadístico SPSS versión 21:

Datos / Seleccionar Casos / Muestra Aleatoria – muestra / Exactamente 22 de 301 / continuar / aceptar.

El esquema de muestreo empleado en cada estrato fue el muestreo aleatorio simple sin reposición.

d) Instrumento de recolección de información Diseño de la Muestra:

El cuestionario utilizado se dividió en 2 partes: En una primera parte se registran los datos generales del Alumno seleccionado (Características Generales), los cuales corresponden a su edad, género, EAP y Ciclo que está estudiando. La segunda parte corresponde a las características específicas relacionado al tema de las preguntas delicadas, en donde se registra la respuesta del entrevistado.

El cuestionario utilizado se muestra en el (Anexo 4).

e) Actividades de Campo

Personal de la encuesta y Materiales del trabajo de campo: Para el trabajo de campo se contó con el ambiente de la Oficina 211, cuyos miembros (las profesoras de esa oficina nos concedieron ese ambiente para llevarse a cabo la encuesta). Un profesional de psicología y la autora de la tesis, fueron los entrevistadores.

Para la difusión de los alumnos seleccionados, se colocó un Comunicado, con la lista de alumnos elegidos según Escuela Académico Profesional, en una pizarra ubicada en la entrada de la FCM para que ellos pudieran enterarse que habían sido elegidos.

La entrevista se realizó en la Oficina 211, en el horario de 10am a 8pm, la cual empezó el día 9 de julio y finalizó el 16 de julio.

3.4. RESULTADOS DE LA APLICACIÓN DEL MODELO DE ENTREVISTA DIRECTA.

3.4.1. Evaluación del rendimiento de la muestra.

Se logró entrevistar a 111 alumnos seleccionados, lo cual significó una tasa de respuesta del 100%.

3.4.2. Características de la población investigada

La población investigada es de 1521 alumnos matriculados el Semestre 2015-I, cuyas edades fluctúan entre 16 y 66 años.

3.4.3. Resultados numéricos.

Utilizando las fórmulas de los estimadores de la proporción de respondientes en la población y las varianzas del estimador de la ecuación (2.4), se obtuvieron los siguientes resultados utilizando el software R-Studio 3.1.2.

Sea $\hat{\pi}_D$ el estimador del modelo de entrevista directa, cuyo estimador es:

$$\hat{\pi}_D = \frac{a}{n}$$

Donde a es el número de entrevistados que respondieron “Sí”. Y la varianza del estimador es la ecuación 2.4

a) Copia en los exámenes:

El estimador del modelo de entrevista directa para la “Proporción de alumnos de la Facultad de Ciencias Matemáticas que han copiado de otros estudiantes durante un examen (práctica) por lo menos una vez”, es:

$$\hat{\pi}_D = 0.514.$$

$$Var(\hat{\pi}_D) = \frac{\pi(1 - \pi)}{n} = 0.00225$$

Y la desviación estándar es

$$\sqrt{Var(\hat{\pi}_D)} = \sqrt{0.00225} = 0.047434$$

Hallando los intervalos de confianza:

$$[\hat{\pi}_{T,inf}, \hat{\pi}_{T,sup}] = [0.435978013, 0.592021987]$$

De aquí, se obtiene la proporción de que los alumnos que hayan copiado de otros estudiantes durante un examen (práctica) por lo menos una vez es: 0.514 (51.4%).

b) Uso de notas en un examen:

Se obtiene la proporción de que los alumnos hayan utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares) es:

$$\hat{\pi}_D = 0.315.$$

$$Var(\hat{\pi}_D) = \frac{\pi(1 - \pi)}{n} = 0.00194$$

Y la desviación estándar es

$$\sqrt{Var(\hat{\pi}_D)} = \sqrt{0.00194} = 0.044045$$

Hallando los intervalos de confianza:

$$[\hat{\pi}_{T,inf}, \hat{\pi}_{T,sup}] = [0.242552422, 0.387447578]$$

c) Consumo de drogas para mejorar el rendimiento académico

La proporción de que los alumnos hayan utilizado drogas para mejorar su rendimiento académico en un examen o práctica por lo menos una vez es:

$$\hat{\pi}_D = 0.009.$$

$$Var(\hat{\pi}_D) = \frac{\pi(1 - \pi)}{n} = 0.00008$$

Y la desviación estándar es

$$\sqrt{Var(\hat{\pi}_D)} = \sqrt{0.00008} = 0.008944$$

Hallando los intervalos de confianza:

$$[\hat{\pi}_{T,inf}, \hat{\pi}_{T,sup}] = [-0.005711571, 0.023711571]$$

d) Copia de un párrafo intencionalmente adoptado del trabajo de otra persona

La proporción de que los alumnos hayan presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona por lo menos una vez es:

$$\hat{\pi}_D = 0.234.$$

$$Var(\hat{\pi}_D) = \frac{\pi(1 - \pi)}{n} = 0.001616$$

Y la desviación estándar es

$$\sqrt{Var(\hat{\pi}_D)} = \sqrt{0.001616} = 0.0401995$$

Hallando los intervalos de confianza:

$$[\hat{\pi}_{T,inf}, \hat{\pi}_{T,sup}] = [0.167877707, 0.300122293]$$

Tabla 3.5: Modelo Triangular de Respuesta No aleatorizada y Método convencional según la estimación del porcentaje de estudiantes que han copiado en un examen por lo menos una vez y su respectivo margen de error

Método Empleado	$\hat{\pi}$	Margen de Error	Eficiencia Relativa
MTRNA	30.86%	7.99%	3.32
Modelo Convencional	51.40%	7.80%	

Tabla 3.6: Modelo Triangular de Respuesta No aleatorizada y Método convencional según la estimación del porcentaje de estudiantes que han utilizado notas en un examen y su respectivo margen de error.

Método Empleado	$\hat{\pi}$	Margen de Error	Eficiencia Relativa
MTRNA	54.73%	5.65%	1.43
Modelo Convencional	31.50%	7.24%	

Tabla 3.7: Modelo Triangular de Respuesta No aleatorizada y Método convencional según la estimación del porcentaje de estudiantes que han utilizado drogas para mejorar su rendimiento académico en un examen o práctica por lo menos una vez y su respectivo error estándar

Método Empleado	$\hat{\pi}$	Margen de Error	Eficiencia Relativa
MTRNA	12.55%	5.78%	3.38
Modelo Convencional	0.9%	1.47%	

Tabla 3.8: Modelo Triangular de Respuesta No aleatorizada y Método convencional según la estimación del porcentaje de estudiantes que han presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona por lo menos una vez y su respectivo error estándar

Método Empleado	$\hat{\pi}$	Margen de Error	Eficiencia Relativa
MTRNA	71.79%	5.04%	1.39
Modelo Convencional	23.40%	6.61%	

3.5. CONCLUSIONES Y RECOMENDACIONES

1. La aplicación del modelo triangular nos permitió poner en práctica una técnica de muestreo que no se había hecho anteriormente en nuestro país, con un resultado eficiente en encuestas con preguntas sensibles.
2. La estimación de “Copia en los exámenes” utilizando el MTRNA es de 30.86% mientras que en el diseño de entrevista directa es de 51.4%.
3. La estimación de “Uso de notas en un examen” utilizando el MTRNA es de 54.73%, mientras que con el diseño de entrevista directa es de 31.5%; obteniéndose mayor porcentaje con el modelo triangular en relación método convencional.
4. La estimación de “Consumo de drogas para aumentar el rendimiento académico” utilizando el MTRNA es de 12.55%, mientras que con el diseño de entrevista directa es de 0.9 %; obteniéndose mayor porcentaje con el modelo triangular en relación método convencional.
5. La estimación de “Copia de párrafo intencionalmente adoptado del trabajo de otra persona” utilizando el MTRNA es de 71.79%, mientras que con el diseño de entrevista directa es de 23.4 %; obteniéndose mayor porcentaje con el modelo triangular en relación método convencional.
6. Se recomienda tener cuidado con las combinaciones utilizadas en cuanto a la eficiencia relativa dado que para un investigador social dispuesto a utilizar el modelo triangular para un tema sensible, en algunas combinaciones estará obligado a entrevistar a un número muy alto frente a un número pequeño del modelo de la entrevista directa.
7. Por lo dicho anteriormente la desventaja principal está en los gastos en lo que respecta a la capacitación de los entrevistadores y el tiempo que requiere explicar al entrevistado el cuestionario.
8. Se recomienda seguir experimentando estos modelos en muestras más grandes y en temas en donde la pregunta sea altamente sensible.

ANEXOS

1. Demostración: Estimación de π utilizando el método de máxima verosimilitud en el modelo triangular.

La función de máxima verosimilitud para el modelo triangular es:

$$L_T(\pi|Y_{Obs}) = \prod_{i=1}^n \{\pi + (1 - \pi)p\}^{y_{i,T}} \{(1 - \pi)(1 - p)\}^{1-y_{i,T}}$$

$$L_T(\pi|Y_{Obs}) = [\pi + (1 - \pi)p]^{\sum y_{i,T}} [(1 - \pi)(1 - p)]^{\sum (1-y_{i,T})}$$

Aplicando la función de logaritmo natural se tiene que:

$$\begin{aligned} \ell_T(\pi|Y_{Obs}) &= \ln[L_T(\pi|Y_{Obs})] \\ &= \sum y_{i,T} \ln[\pi + (1 - \pi)p] + \left(n - \sum y_{i,T}\right) \ln[(1 - \pi)(1 - p)] \end{aligned}$$

Derivando con respecto al parámetro π e igualando a cero, tenemos:

$$\frac{\partial \ell_T(\pi|Y_{Obs})}{\partial \pi} = \sum y_{i,T} \frac{(1 - p)}{\pi + (1 - \pi)p} - \left(n - \sum y_{i,T}\right) \frac{(1 - p)}{(1 - \pi)(1 - p)}$$

$$\frac{\partial \ell_T(\pi|Y_{Obs})}{\partial \pi} = \frac{\sum y_{i,T} [(1 - p)(1 - \pi)(1 - p)] - (n - \sum y_{i,T})(1 - p)[\pi + (1 - \pi)p]}{[\pi + (1 - \pi)p][(1 - \pi)]}$$

$$\frac{\partial \ell_T(\pi|Y_{Obs})}{\partial \pi} =$$

$$\frac{\sum y_{i,T} [(1 - p)^2(1 - \pi)] - n(1 - p)(p + \pi(1 - p)) + \sum y_{i,T} (1 - p)[p + \pi(1 - p)]}{[\pi + (1 - \pi)p][(1 - \pi)]}$$

$$\frac{\partial \ell_T(\pi|Y_{Obs})}{\partial \pi} = \frac{\sum y_{i,T} [(1 - p)(1 - \pi)] - n(p + \pi(1 - p)) + \sum y_{i,T} [p + \pi(1 - p)]}{[\pi + (1 - \pi)p][(1 - \pi)]}$$

$$\frac{\partial \ell_T(\pi|Y_{Obs})}{\partial \pi} = \frac{\sum y_{i,T} [1 - p - \pi + p\pi + p + \pi - \pi p] - n(p + \pi(1 - p))}{[\pi + (1 - \pi)p][(1 - \pi)]}$$

$$\frac{\partial \ell_T(\pi|Y_{Obs})}{\partial \pi} = \frac{\sum y_{i,T} - n(p + \pi(1 - p))}{[\pi + (1 - \pi)p][(1 - \pi)]} = 0, \text{ si y solo si}$$

$$\frac{\partial \ell_T(\pi|Y_{Obs})}{\partial \pi} = \sum y_{i,T} - n(p + \pi(1 - p)) = 0$$

$$\frac{1}{n} \sum y_{i,T} = p + \pi(1 - p)$$

Y despejando π se tiene como

$$\hat{\pi}_T = \frac{\bar{y}_T - p}{(1 - p)}$$

L.Q.Q.D

2. Resultados de estimaciones con el software R- Studio para el modelo triangular.

Código para Modelo Triangular de Respuesta No aleatorizada

Rosa Ramirez Carrillo

3 de septiembre de 2015

El siguiente código es un formato simple para estimar la proporción de pertenecer al grupo con característica sensible, será trabajado de acuerdo al modelo triangular de respuesta no aleatorizada.

Primer Paso: Ingresar las probabilidades de la característica no sensible (recordar que esta será conocida)

```
p1=5/12  
p2=0.19  
p3=0.23  
p4=0.22
```

Segundo Paso: Importar la data a trabajar

```
data<-read.csv("F:/Nueva carpeta/2015/2015-II/Seminario/Respuesta No Aleatorizada Triangular/Tesis.csv", header=T)
```

Tercer Paso: Seleccionar las preguntas de interés y hallar sus proporciones.

```
pregunta5=data[,5]  
pregunta6=data[,6]  
pregunta7=data[,7]  
pregunta8=data[,8]  
  
y1=mean(pregunta5)  
y2=mean(pregunta6)  
y3=mean(pregunta7)  
y4=mean(pregunta8)
```

Cuarto Paso: Hallar probabilidad de presentar la característica sensible:

```
Pregunta_1=(y1-p1)/(1-p1)  
Pregunta_2=(y2-p2)/(1-p2)  
Pregunta_3=(y3-p3)/(1-p3)  
Pregunta_4=(y4-p4)/(1-p4)
```

Pregunta 1: Que porcentaje de alumnos copian de otros estudiantes durante un examen (práctica).

```
100*Pregunta_1  
## [1] 30.85714
```


Pregunta 2: Que porcentaje de alumnos ha utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares)

```
100*Pregunta_2
```

```
## [1] 54.73251
```

Pregunta 3: Que porcentaje de alumnos ha utilizado drogas para mejorar su rendimiento académico en un examen o práctica.

```
100*Pregunta_3
```

```
## [1] 12.55411
```

Pregunta 4: Que porcentaje de alumnos ha presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona.

```
100*Pregunta_4
```

```
## [1] 71.79487
```

Ahora conocemos la probabilidad de pertenecer a los grupos con característica sensible.

Hallando la Varianza y la desviación estándar en cada variable:

```
Varianza_1=(Pregunta_1*(1-Pregunta_1)/300)+(p1*(1-Pregunta_1)/(300*(1-p1)))
```

```
Varianza_1
```

```
## [1] 0.002357442
```

```
desv_est1=(Varianza_1)^(0.5)
```

```
desv_est1
```

```
## [1] 0.0485535
```

```
Varianza_2=(Pregunta_2
```

```
*(1-Pregunta_2)/300)+(p2*(1-Pregunta_2)/(300*(1-p2)))
```

```
Varianza_2
```

```
## [1] 0.001179811
```

```
desv_est2=(Varianza_2)^(0.5)
```

```
desv_est2
```

```
## [1] 0.03434838
```

```
Varianza_3=(Pregunta_3*(1-Pregunta_3)/300)+(p3*(1-Pregunta_3)/(300*(1-p3)))
```

```
Varianza_3
```

```
## [1] 0.001236609
```

```
desv_est3=(Varianza_3)^(0.5)
```

```
desv_est3
```

```
## [1] 0.03516544

Varianza_4=(Pregunta_4*(1-Pregunta_4)/300)+(p4*(1-Pregunta_4)/(300*(1-p4)
))
Varianza_4

## [1] 0.0009401709

desv_est4=(Varianza_4)^(0.5)
desv_est4

## [1] 0.03066221
```

3. Cuestionario 1: Modelo Triangular de Respuesta No Aleatorizada

UNMSM

Facultad de Ciencias Matemáticas

Estimado alumno, usted ha sido elegido al azar para participar en una investigación sobre algunas características de la conducta de los estudiantes de la FCM. Agradecemos de antemano su valiosa colaboración.

CUESTIONARIO SOBRE ALGUNAS CARACTERÍSTICAS DE LOS ALUMNOS

I. CARACTERÍSTICAS GENERALES

1) Edad: _____ 2) Género: 1. Femenino 2. Masculino

3) E.A.P.: _____ 4) CICLO: _____

II. CARACTERÍSTICAS DEL TEMA

Por favor responda con veracidad, marque una cruz en el círculo superior si Ud. Pertenece al círculo o marque una cruz en el cuadrado superior si Ud. Pertenece a uno de los 3 cuadrados.

5)

	No nació entre Agosto y Diciembre. (W = 0)	Nació entre Agosto y Diciembre (W = 1)
Categoría		

Nunca he copiado de
otros estudiantes
durante un examen
(práctica).

(Y = 0)

☐
☐

He copiado de otros
estudiantes durante un
examen (práctica) por
lo menos una vez.

(Y = 1)

☐
☐

6)

Categoría	Nunca interrumpí mis estudios universitarios (W = 0)	He interrumpido mis estudios universitarios por lo menos una vez (W = 1)
Nunca he utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares) (Y = 0)	<input type="radio"/>	<input type="checkbox"/>
He utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares) (Y = 1)	<input type="checkbox"/>	<input type="checkbox"/>

7)

Categoría	No ingrese por la Modalidad de Centro Preuniversitario (W = 0)	Ingrese por la Modalidad de Centro Pre Universitario (W = 1)
Nunca he utilizado drogas para mejorar mi rendimiento académico en un examen o práctica (Y = 0)	<input type="radio"/>	<input type="checkbox"/>
He utilizado drogas para mejorar mi rendimiento académico en un examen o práctica por lo menos una vez (Y = 1)	<input type="checkbox"/>	<input type="checkbox"/>

8)

Categoría	No me gusta estudiar solo(a) (W = 0)	Me gusta estudiar solo (a) (W = 1)
Nunca he presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona. (Y = 0)	<input type="radio"/>	<input type="checkbox"/>
He presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona por lo menos una vez. (Y = 1)	<input type="checkbox"/>	<input type="checkbox"/>

4. Cuestionario 2: Método de Entrevista Directa

UNMSM
Facultad de Ciencias Matemáticas

CUESTIONARIO SOBRE EL COMPORTAMIENTO DE LOS ALUMNOS DE LA FCM FRENTE A PREGUNTAS SENSIBLES

N° de Cuestionario

--	--	--

I. CARACTERÍSTICAS GENERALES

1. Edad: _____

2. Género

1. Femenino

2. Masculino

3. E.A.P: _____

4. CICLO:

1) I 2) II 3) III 4) IV 5) V 6) VI 7) VII 8) VIII 9) IX 10) X

II. CARACTERÍSTICAS DEL TEMA

5. ¿Has copiado de otros estudiantes durante un examen (práctica) por lo menos una vez?

1. Si

2. No

6. ¿Has utilizado notas en un examen (práctica) (incluyendo notas en los teléfonos móviles, calculadoras u otras similares)?

1. Si

2. No

7. ¿Has utilizado drogas para mejorar tu rendimiento académico en un examen o práctica por lo menos una vez?

1. Si

2. No

8. ¿Has presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona por lo menos una vez?

1. Si

2. No

Nombre del entrevistador: _____

Fecha: _____

5. Resultados de la Encuesta del Modelo Triangular

Orden	E.A.P.	Pregunta 5		Pregunta 6		Pregunta 7		Pregunta 8	
		Círculo = 0	Cuadrado= 1	Círculo = 0	Cuadrado= 1	Círculo = 0	Cuadrado= 1	Círculo = 0	Cuadrado= 1
1	computación	0		0		0		0	
2	computación	1		0		0		0	
3	computación	1		1		0		1	
4	computación	0		0		0		1	
5	computación	0		1		0		1	
6	computación	0		1		0		0	
7	computación	1		1		1		1	
8	computación	1		1		1		1	
9	computación	0		0		1		1	
10	computación	0		1		0		1	
11	computación	0		1		1		1	
12	computación	1		1		1		1	
13	computación	1		1		1		1	
14	computación	1		0		1		1	
15	computación	1		1		1		1	
16	computación	1		1		0		1	
17	computación	1		1		0		0	
18	computación	1		0		0		1	
19	computación	1		1		1		1	
20	computación	1		1		1		1	
21	computación	0		1		1		1	
22	computación	0		0		0		0	
23	computación	0		1		1		1	
24	computación	1		1		1		1	
25	computación	1		1		1		1	
26	computación	1		1		1		1	
27	computación	1		0		0		1	
28	computación	0		1		0		0	
29	computación	1		0		0		1	
30	computación	0		0		0		0	
31	computación	1		1		0		1	
32	computación	1		1		0		1	

33	computación	0	1	0	1
34	computación	1	1	0	0
35	computación	0	1	1	1
36	computación	0	0	1	1
37	computación	0	1	0	0
38	computación	0	1	1	1
39	computación	1	1	0	1
40	computación	1	1	0	1
41	computación	0	1	0	0
42	computación	0	1	0	1
43	computación	1	1	0	0
44	computación	1	1	1	1
45	computación	0	0	0	1
46	computación	1	1	0	1
47	computación	1	1	1	1
48	computación	1	1	1	1
49	computación	0	0	1	0
50	computación	1	1	0	0
51	computación	1	0	0	0
52	computación	1	1	0	0
53	computación	0	0	1	1
54	computación	1	0	1	1
55	computación	0	0	0	1
56	computación	0	1	0	1
57	computación	1	1	1	0
58	computación	0	0	0	1
59	computación	1	1	0	1
60	Estadística	0	0	0	1
61	Estadística	0	0	0	1
62	Estadística	1	1	0	1
63	Estadística	1	0	0	1
64	Estadística	1	1	0	1
65	Estadística	0	0	0	1
66	Estadística	0	1	0	1
67	Estadística	0	1	0	1
68	Estadística	0	1	0	1
69	Estadística	1	1	0	1
70	Estadística	1	1	1	1
71	Estadística	0	0	0	1
72	Estadística	1	0	0	0

73	Estadística	0	1	0	1
74	Estadística	1	1	0	0
75	Estadística	0	1	0	1
76	Estadística	0	0	0	1
77	Estadística	1	0	0	1
78	Estadística	1	0	1	0
79	Estadística	0	0	0	1
80	Estadística	1	1	0	1
81	Estadística	1	1	1	1
82	Estadística	1	1	0	1
83	Estadística	0	0	0	1
84	Estadística	1	0	0	1
85	Estadística	0	1	0	1
86	Estadística	0	0	1	0
87	Estadística	0	1	0	1
88	Estadística	1	1	0	1
89	Estadística	1	1	0	1
90	Estadística	1	1	0	1
91	Estadística	1	1	0	1
92	Estadística	0	1	1	1
93	Estadística	1	1	0	1
94	Estadística	1	1	0	1
95	Estadística	1	1	0	1
96	Estadística	1	0	0	1
97	Estadística	1	1	0	1
98	Estadística	0	0	0	1
99	Estadística	1	1	0	1
100	Estadística	0	1	0	1
101	Estadística	1	1	0	1
102	Estadística	1	1	0	1
103	Estadística	1	0	0	0
104	Estadística	0	0	0	1
105	Estadística	1	0	1	1
106	Estadística	0	0	0	0
107	Estadística	1	1	0	1
108	Estadística	1	1	0	0
109	Estadística	0	1	0	1
110	Estadística	1	1	0	1
111	Estadística	1	1	0	1
112	Estadística	1	1	1	1

113	Estadística	1	1	0	1
114	Estadística	1	1	0	1
115	Estadística	1	1	0	1
116	Estadística	0	0	0	1
117	Estadística	1	1	0	1
118	Inv.operativa	0	1	0	1
119	Inv.operativa	0	1	1	1
120	Inv.operativa	1	1	1	1
121	Inv.operativa	1	1	1	1
122	Inv.operativa	1	1	1	1
123	Inv.operativa	0	0	1	1
124	Inv.operativa	1	1	1	1
125	Inv.operativa	1	1	1	1
126	Inv.operativa	1	1	0	0
127	Inv.operativa	1	1	0	0
128	Inv.operativa	0	0	1	0
129	Inv.operativa	0	0	1	1
130	Inv.operativa	1	1	0	1
131	Inv.operativa	0	0	0	0
132	Inv.operativa	1	1	1	0
133	Inv.operativa	0	1	1	0
134	Inv.operativa	1	1	1	1
135	Inv.operativa	1	1	1	1
136	Inv.operativa	1	1	1	1
137	Inv.operativa	0	1	1	1
138	Inv.operativa	1	1	0	0
139	Inv.operativa	0	0	0	1
140	Inv.operativa	1	1	1	1
141	Inv.operativa	1	1	0	1
142	Inv.operativa	0	0	1	1
143	Inv.operativa	1	1	1	1
144	Inv.operativa	1	1	0	1
145	Inv.operativa	1	1	0	1
146	Inv.operativa	1	1	0	0
147	Inv.operativa	0	1	1	1
148	Inv.operativa	1	1	1	1
149	Inv.operativa	0	1	0	1
150	Inv.operativa	0	1	0	1
151	Inv.operativa	0	0	0	1
152	Inv.operativa	1	0	1	1

153	Inv.operativa	1	1	1	1
154	Inv.operativa	0	1	1	1
155	Inv.operativa	1	0	0	0
156	Inv.operativa	0	0	1	0
157	Inv.operativa	1	1	0	0
158	Inv.operativa	0	0	1	1
159	Inv.operativa	1	1	0	0
160	Inv.operativa	1	0	0	0
161	Inv.operativa	1	0	0	1
162	Inv.operativa	0	0	0	1
163	Inv.operativa	1	0	0	1
164	Inv.operativa	0	1	1	1
165	Inv.operativa	1	1	1	1
166	Inv.operativa	1	1	1	1
167	Inv.operativa	0	0	0	1
168	Inv.operativa	1	1	1	1
169	Inv.operativa	1	1	0	1
170	Inv.operativa	0	0	0	1
171	Inv.operativa	1	1	0	1
172	Inv.operativa	0	1	0	1
173	Inv.operativa	1	1	0	0
174	Inv.operativa	0	0	0	0
175	Inv.operativa	1	1	0	1
176	Inv.operativa	0	1	0	0
177	Inv.operativa	1	1	1	1
178	Inv.operativa	1	1	0	0
179	Inv.operativa	0	0	1	1
180	Inv.operativa	1	0	0	1
181	Inv.operativa	0	0	0	1
182	Inv.operativa	1	1	0	1
183	Inv.operativa	0	1	0	0
184	Inv.operativa	0	1	0	0
185	Inv.operativa	0	0	0	1
186	Inv.operativa	1	1	1	0
187	Inv.operativa	0	1	0	0
188	Inv.operativa	0	1	0	1
189	Inv.operativa	1	1	1	0
190	Inv.operativa	1	1	1	0
191	Inv.operativa	0	1	0	0
192	Inv.operativa	1	1	0	1

193	Inv.operativa	0	1	0	0
194	Inv.operativa	1	1	1	1
195	Inv.operativa	1	1	1	1
196	Inv.operativa	1	0	0	0
197	Inv.operativa	1	1	0	1
198	Inv.operativa	1	0	1	1
199	Inv.operativa	0	0	0	0
200	Inv.operativa	1	1	0	0
201	Inv.operativa	1	1	0	1
202	Inv.operativa	1	1	0	1
203	Inv.operativa	1	1	0	1
204	Inv.operativa	1	1	0	1
205	Inv.operativa	1	0	0	1
206	Inv.operativa	0	0	0	1
207	Inv.operativa	1	0	0	1
208	Inv.operativa	1	1	0	1
209	Inv.operativa	0	0	0	1
210	Inv.operativa	1	1	0	1
211	Matematica	0	1	0	1
212	Matematica	0	0	0	1
213	Matematica	1	0	1	0
214	Matematica	1	1	0	1
215	Matematica	1	1	0	1
216	Matematica	0	0	0	1
217	Matematica	1	1	0	0
218	Matematica	0	1	0	0
219	Matematica	0	1	0	0
220	Matematica	0	0	0	1
221	Matematica	1	0	1	0
222	Matematica	1	0	0	1
223	Matematica	0	1	0	0
224	Matematica	0	1	0	1
225	Matematica	0	1	1	1
226	Matematica	1	1	0	1
227	Matematica	0	0	1	1
228	Matematica	1	1	1	1
229	Matematica	1	1	1	1
230	Matematica	0	1	1	1
231	Matematica	1	1	0	1
232	Matematica	1	0	0	1

233	Matematica	0	0	0	1
234	Matematica	1	1	0	1
235	Matematica	0	0	0	0
236	Matematica	0	0	1	1
237	Matematica	1	0	0	1
238	Matematica	1	0	1	1
239	Matematica	0	0	1	1
240	Matematica	1	1	0	1
241	Matematica	1	1	1	0
242	Matematica	0	0	0	0
243	Matematica	0	0	0	1
244	Matematica	0	0	0	1
245	Matematica	1	0	1	1
246	Matematica	1	1	1	1
247	Matematica	1	1	1	1
248	Matematica	0	1	0	1
249	Matematica	0	1	1	1
250	Matematica	1	0	0	1
251	Matematica	1	1	0	1
252	Matematica	0	1	1	1
253	Matematica	0	0	0	1
254	Matematica	1	0	0	0
255	Matematica	1	0	0	0
256	Matematica	1	1	0	1
257	Matematica	0	0	0	0
258	Matematica	0	0	0	0
259	Matematica	1	0	0	1
260	Matematica	0	1	0	1
261	Matematica	1	0	0	1
262	Matematica	1	0	1	1
263	Matematica	1	1	0	1
264	Matematica	0	1	0	1
265	Matematica	1	1	1	1
266	Matematica	1	0	0	1
267	Matematica	0	1	0	1
268	Matematica	1	0	1	1
269	Matematica	1	1	0	1
270	Matematica	1	1	0	1
271	Matematica	1	0	0	1
272	Matematica	1	1	1	1

273	Matematica	0	1	1	1
274	Matematica	0	1	1	1
275	Matematica	1	1	1	1
276	Matematica	0	1	1	1
277	Matematica	1	1	0	1
278	Matematica	0	0	0	1
279	Matematica	1	0	0	1
280	Matematica	1	1	0	1
281	Matematica	0	1	0	1
282	Matematica	1	1	0	1
283	Matematica	1	0	1	0
284	Matematica	0	0	0	1
285	Matematica	1	0	0	1
286	Matematica	0	0	0	0
287	Matematica	1	1	0	1
288	Matematica	1	1	1	1
289	Matematica	0	1	0	1
290	Matematica	1	1	0	1
291	Matematica	0	0	0	1
292	Matematica	1	0	0	1
293	Matematica	1	1	1	1
294	Matematica	1	1	0	1
295	Matematica	1	0	0	1
296	Matematica	1	0	0	1
297	Matematica	1	1	1	1
298	Matematica	0	0	0	1
299	Matematica	0	0	0	1
300	Matematica	1	0	1	1

Observación:

- Pregunta 5: ¿Has copiado de otros estudiantes durante un examen (práctica).?
- Pregunta 6: ¿Has utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares).?
- Pregunta 7: ¿Has utilizado drogas para mejorar mi rendimiento académico en un examen o práctica?
- Pregunta 8: ¿Has presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona?

6. Resultados de la Encuesta por Entrevista Directa

Orden	E.A.P.	Pregunta 1		Pregunta 2		Pregunta 3		Pregunta 4	
		No = 0	Si = 1	No = 0	Si = 1	No = 0	Si = 1	No = 0	Si = 1
1	Comp.Cientifica	1		1		0		0	
2	Comp.Cientifica	0		0		0		0	
3	Comp.Cientifica	1		0		0		0	
4	Comp.Cientifica	0		0		0		0	
5	Comp.Cientifica	1		1		0		0	
6	Comp.Cientifica	0		0		0		0	
7	Comp.Cientifica	1		1		0		0	
8	Comp.Cientifica	1		1		0		0	
9	Comp.Cientifica	1		0		0		1	
10	Comp.Cientifica	0		0		0		0	
11	Comp.Cientifica	0		0		0		1	
12	Comp.Cientifica	1		1		0		0	
13	Comp.Cientifica	0		1		0		1	
14	Comp.Cientifica	1		0		0		0	
15	Comp.Cientifica	0		0		0		0	
16	Comp.Cientifica	1		0		0		0	
17	Comp.Cientifica	1		1		0		0	
18	Comp.Cientifica	0		0		0		0	
19	Comp.Cientifica	0		1		0		0	
20	Comp.Cientifica	0		0		0		0	
21	Comp.Cientifica	0		1		0		0	
22	Comp.Cientifica	0		0		0		0	
23	Estadística	1		1		0		0	
24	Estadística	1		1		0		0	
25	Estadística	1		1		0		1	
26	Estadística	0		0		0		1	
27	Estadística	1		0		0		0	
28	Estadística	1		1		0		0	
29	Estadística	1		1		0		0	
30	Estadística	1		1		0		0	
31	Estadística	1		1		0		1	
32	Estadística	1		0		0		0	
33	Estadística	1		0		0		0	
34	Estadística	1		1		0		1	

35	Estadística	0	1	0	0
36	Estadística	0	0	0	0
37	Estadística	1	1	0	0
38	Estadística	1	0	0	0
39	Estadística	0	0	0	0
40	Estadística	0	1	0	0
41	Estadística	1	1	0	1
42	Estadística	1	0	0	1
43	Estadística	1	1	0	1
44	Inv.operativa	0	0	0	0
45	Inv.operativa	1	1	0	0
46	Inv.operativa	1	0	0	1
47	Inv.operativa	1	0	0	0
48	Inv.operativa	1	1	0	0
49	Inv.operativa	0	0	0	0
50	Inv.operativa	0	0	0	0
51	Inv.operativa	1	0	0	0
52	Inv.operativa	0	1	0	0
53	Inv.operativa	0	0	0	0
54	Inv.operativa	0	0	0	0
55	Inv.operativa	1	0	0	1
56	Inv.operativa	1	1	0	0
57	Inv.operativa	1	0	0	0
58	Inv.operativa	1	1	0	1
59	Inv.operativa	0	0	0	0
60	Inv.operativa	1	1	0	0
61	Inv.operativa	0	0	0	1
62	Inv.operativa	1	0	0	0
63	Inv.operativa	1	0	0	1
64	Inv.operativa	0	0	0	1
65	Inv.operativa	0	0	0	0
66	Inv.operativa	0	0	0	0
67	Inv.operativa	0	0	0	0
68	Inv.operativa	0	0	0	0
69	Inv.operativa	0	0	0	0
70	Inv.operativa	0	0	0	0
71	Inv.operativa	1	0	0	0
72	Inv.operativa	1	0	0	1
73	Inv.operativa	1	0	0	1
74	Inv.operativa	0	0	0	1

75	Inv.operativa	1	0	0	1
76	Inv.operativa	0	1	0	0
77	Inv.operativa	1	0	0	0
78	Matematica	1	0	0	0
79	Matematica	0	0	0	0
80	Matematica	0	0	0	1
81	Matematica	0	1	0	0
82	Matematica	1	0	1	1
83	Matematica	0	0	0	0
84	Matematica	1	0	0	1
85	Matematica	1	0	0	0
86	Matematica	1	0	0	0
87	Matematica	0	0	0	0
88	Matematica	0	0	0	0
89	Matematica	1	0	0	1
90	Matematica	1	0	0	0
91	Matematica	0	0	0	0
92	Matematica	0	0	0	0
93	Matematica	0	0	0	0
94	Matematica	0	0	0	0
95	Matematica	0	1	0	0
96	Matematica	0	1	0	0
97	Matematica	1	0	0	0
98	Matematica	1	0	0	1
99	Matematica	0	0	0	0
100	Matematica	1	1	0	0
101	Matematica	0	0	0	0
102	Matematica	1	0	0	0
103	Matematica	0	0	0	0
104	Matematica	0	0	0	0
105	Matematica	1	0	0	0
106	Matematica	0	0	0	0
107	Matematica	1	1	0	0
108	Matematica	0	0	0	0
109	Matematica	0	0	0	1
110	Matematica	0	0	0	0
111	Matematica	1	1	0	0

Observación:

- Pregunta 1: ¿Has copiado de otros estudiantes durante un examen (práctica).?
- Pregunta 2: ¿Has utilizado notas en un examen (incluyendo notas en los teléfonos móviles, calculadoras u otras similares).?
- Pregunta 3: ¿Has utilizado drogas para mejorar mi rendimiento académico en un examen o práctica?
- Pregunta 4: ¿Has presentado un documento que contiene un párrafo intencionalmente adoptado del trabajo de otra persona?

REFERENCIAS BIBLIOGRÁFICAS

- Basulto, J. (1982). El diseño de respuesta aleatorizada de warner. Un modelo de superpoblación., (96), 51 a 62.
- Cochran, W. (1977). \iSampling Techniques. New York: Jhon Wiley & Sons. (Third edition).
http://hbanaszak.mjr.uw.edu.pl/TempTxt/Cochran_1977_Sampling_Techniques__Third_Edition.pdf
- Guo-Liang Tian, J.-W. Y. (2007). A new non-randomized model for analysis sensitive questions with binary outcomes. *Statistics in medicine*, 26(23), 4238-52.
<http://doi.org/10.1002/sim.2863>
- Nayak, T. K. (1994). On randomized response surveys for estimating a proportion. *Communications in Statistics - Theory and Methods*, 23(11), 3303-3321.
<http://doi.org/10.1080/03610929408831448>
- Shein-Chung Chow, Hansheng Wang, & Jun Shao. (2003). *Sample Size Calculations in Clinical Research*.
- Tan, M. T., Tian, G.-L., & Tang, M.-L. (2009). Sample Surveys With Sensitive Questions: A Nonrandomized Response Approach. *The American Statistician*, 63(1). Recuperado a partir de
<http://search.proquest.com/docview/228530647/4D36C03AECB64134PQ/1>
- Tian, G.-L., & Tang, M.-L. (2014). *Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys*. CRC Press.
- Tian, G.-L., Tang, M.-L., Liu, Z., Tan, M., & Tang, N.-S. (2011). Sample size determination for the non-randomised triangular model for sensitive questions in

a survey. *Statistical Methods in Medical Research*, 20(3), 159-73.

<http://doi.org/http://dx.doi.org/10.1177/0962280208099444>

Tracy PE, & Fox JA. (1986). Randomized response: a method for sensitive surveys.

Recuperado a partir de <http://www.popline.org/node/345751>

Vakilian, K., Mousavi, S. A., & Keramat, A. (2014). Estimation of sexual behavior in the 18-to-24-years-old Iranian youth based on a crosswise model study. *BMC Research Notes*, 7, 28. <http://doi.org/http://dx.doi.org/10.1186/1756-0500-7-28>

Yu, J.-W., Tian, G.-L., & Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika*, 67(3), 251-263.

<http://doi.org/10.1007/s00184-007-0131-x>